Co-China 周刊

NO. 160
2014年 6月27日

韩彦文：超越逻辑 | 黄永明：模拟人脑，迄今最大规模 |
陈赛：黑镜中照见自己的脸

人工智能之惑





编者的话

继大数据之后，月末版带领大家一窥几十年前科技领域就兴起的一波热潮：人工智能。尽管不是“可疑”的人工智能，但人类从这一门学科或技术的伊始到现在经历了热切的期待，继而失望，再到如今更为谨慎的态度。

Gary Morgenthaler 认为，不论态度乐观还是悲观，人工智能的时代都已经来临。因此对于商业世界，更明智的态度或许应是选择一条正确聪明的应用人工智能的策略。随后的两篇文章分别提供了人工智能在音乐和学习领域的实际应用，但同时也抛出了疑问：如果机器可以像莫扎特一样谱曲或如人工神经网络一般进行深度学习，那么作为生物的人还有什么独特之处？凯文·凯利则认为，机器的出现，并不是为了成为人类的敌人，而是变成与人类并肩作战的战友。

人类常常质疑人工智能的未来，源于其原理。韩彦文指出，人工智能无法超越人脑本身的逻辑，因为计算机的工作不过是将最简单的逻辑不断重复，而人脑可以一蹴而就完成复杂的分析过程。即使模拟人脑在人脑面前也无法骄傲，因为前者依旧无法思考或感受。《经济学人》关于昆虫智慧的文章认为，昆虫的跳跳舞恰恰是人工智能的核心——后者具有的认知功能，正因为神经元形似“跳跳舞”的互动。哲学家希尔勒更追溯根源，指出人工智能一词的谬误：这意味着“智能”既可以是实际可以生产出来的产品，也可以是人类智慧的仿造品。

人工智能更大的意义在于它更深刻地向人类启示“人何以为人”。 即使获得了 2008 年勒布纳人工智能奖，获奖计算机程序 Elbot 也并不是比以往计算机更加智能，不过因为人不再智能。Gary Marcus 点出，人工智能的发展可能超过人类的控制，因为后者并不清楚自己在创造什么东西。这也是许多人迷上电视剧《黑镜》的原因之一，我们热爱人工智能，甚至在不知不觉成为它的奴隶，并且心甘情愿。关于人工智能最可悲的恐怕是，我们对它依赖，以至于情感的需求也可以通过物化的人工智能得到满足。

Co-China 周刊由「我在中国」（Co-China）论坛志愿者团队制作，每周出版一期，周刊通过网络发布，所有非一五一十部落的文章均经过作者或首发媒体的授权，期待大家的关注和建议。



目录














【应用·人工智能】

Gary Morgenthaler：人工智能时代已经来临！

Gary Morgenthaler， 风投机构 Morgenthaler Ventures 的合伙人之一



乐观主义者预见了人机和谐的人工智能乌托邦；悲观主义者看到类似《终结者3》中天网的反乌托邦的未来。但是现在，很明确的是，主流AI的时代正在到来。商人和投资者的或许更应该关心的是，“你的AI策略是什么？

”

经历数十年的失败，受益于计算能力的巨变，人工智能终于可以面对公用。

50年的不断失败和次要的成功， 人工智能（Artificial Intelligence, AI）正在走向主流。一些趋势的汇集——云计算，智能手机，拓展的带宽容量，改进的智能算法，以及根据摩尔法则不断增强的处理能力极大的加速了人工智力的发展。人工智能的产业下属项目——语音识别，自然语言理解，机器的学习，计算机视觉等等不仅都有所发展，还开始了合作。我们，终于，即将见证人类智力的精妙。在这个过程中，我们正从“刚刚好”走向“出乎意料”！

在接下来的三到五年内，你将能够：

1. 使用虚拟个人助手（Virtual Personal Assistant，VPA）来管理你的商业或者社交日程表。 VPA将会轻松的为下周的四人会议找到一个合适的时间，或者为你在俱乐部预定一个下午两点的下午茶。他们会了解你的社交图表以及作息规律，而且像任何好的经理助手一样，提出不错的建议。（“你想让我把你的会计师也邀请到会议里吗？”）

2. 直接向现在的“Roomba”的后裔产品发消息说“给楼下吸尘”，然后一点也不担心楼下的猫和古董桌。

3. 放纵的享用只有正宗日本料理才能提供的台端烧和BBQ。你将能够找出距你居住的大阪酒店最近的这样的餐馆，判断它的营业时间，然后通过你智能手机的翻译软件来预定一个位子。

4. 向搜索引擎提出关于事实的问题，然后得到明确的回答，而不是一百万个蓝色的超链接。不用再说那些模糊的语言了，比如“瑞奇-韩德森”，“队长”，“上垒”，“盗垒”，相反，你只要问“瑞奇-韩德森在哪年带领美国队上垒和盗垒？”(答案：1982，1983，1989 和 1998)

于此同时，竞争激烈的 IT 业也会产生深远的变革：

1. 大多数智能手机将会通过语音启动而不是拇指。顾客会要求通过语音来启动搜索，短信，邮件，日程表，协作和购买等服务。

2. 人工智能将会成为智能手机竞争的关键，甚至全 IT 行业领导权的关键。

要想理解这个巨变的内幕需要回顾一下人工智能的曲折历史。人工智能在 1956 年被设立为一个项目，之后带给人们极高的期待（项目的确立者之一，Herb Simon 曾说，“在 20 年内，电脑就会做到人类能做的任何事”）和痛苦的失望。即便数十年以来实验室不断传出令人激动的演示，AI 总是让人觉得在现实中不靠谱。

那 AI 的自然语言理解来举个例子。比如这句，“我想出车”是说话者想要开车出去，还是下车出去？（这部分有些改动——译者）大多数人能够轻易明白的意思，已证实会让 AI 难以理解。尽管有更高级的自然语言算法的相助，但是终极的解决方法总是在几百万个口语样式和几乎同样多的说话人中迭代。对于 AI 来说，现实的适用性依靠的是强大计算的蛮力。

AI 显著的提高需要等待摩尔定律预示的更强大计算能力的到来。在上世纪 90 年代末，处理能力已经达到了这样一个高度：有了合适的算法，计算机就可以可靠地识别用来预定机票和股票转账几百个基本词汇。同语音识别一起，人工智能已经在其他相对较小的领域展示出了商业成功，比如逻辑，数据挖掘和医学诊断。

云的降临

AI 的最近进展是这样的：逐渐的，不完全的成功被隔离在彼此很少沟通的各个仓筒里。而现在进入到了云计算时代，尽管这个技术主要优势在于数据管理，而且也不是很新，但是 AI 开发者还是在最近几年对其大加利用。他们找到了与仓筒中的 AI 信息协同工作的新方法并且，重建了一个更拓展的 AI 学科。

首先，创建更够容易获取的数据，云计算设备数据以及众包的数据池。这些技术可以智能的整理百万条人们普遍了解的“常识”。比如真知公司（True Knowledge）自动的利用维基百科，自由基和其他数据库作为源来分析“事实”，那些“事实”都是由志愿者不辞劳苦的整理输入的。基于 AI 的常识理性，尽管还没有完成，但是可谓近在咫尺了。



其次，那些数据池设备快速的反馈。机器学习算法能够自动的识别复杂的样式然后根据数据做出智能的决定，这个做法可以应用到大型的数据集上，并利用经验不断的提高性能。

令人疲惫的旧方法

那是一个里程碑式的进展。十年前， Nuance 公司（一个语音识别公司，我对其投资并在其董事会上供职），需要人工在广泛的人口中收集分离的语音回应平台。举个例子，公司需要捕获波士顿的人和田纳西州乡村地区的人的“火”一词的发音。而且，Nuance 的工程师将会每隔一段时间的把那个信息向下一次软件发布（每隔 18 个月）的数据集中合并。

而现今，Nuance 和其他语音识别公司，比如 Vlingo，只需在数据池中抽取一些信息就可以和数以百万记的发音比较并把结果几乎是实时的传回系统。只有一个结果：一个可以获取的商业软件，它可以以 100%的准确率识别英语中的数万单词。其他的人工智能比较机器，比如面部识别，也可以以同样的机理增强他们的产品。

最后，也是最重要的一点，云计算可以加快摩尔定律所预测的处理速度并且让 AI 的应用狭窄的子科目协同合作。从 1990 年到 2020 年，摩尔定律本身就可提高一个服务器的处理性能一百万倍。在 2020 年，对比之下，云计算可以让与人工智能相关的处理能力乘以十亿。云计算和开源程序（比如由 Hadoop 开发的 YAHOO！）的结合，可以让 AI 系统在多个服务器上同时的运行数据和算法。对于 AI 来说，那意味着我们可以摈弃那些笨重的仓筒子科目（机器学习，语音识别，对话管理等等），而是同时处理他们并且把他们有机的整合到一起。

准备好应对仿生群

在未来，电脑群，像蚂蚁的殖民地或者欧椋鸟群一样，将会被云计算带领起来解决全球性问题。（如果你听起来像威廉•吉布森 1984 年的科幻小说，《神经漫游者》中的仿生群，那就对了）。人工智能研发者已经给这种现象定义了一个官方的名字：群体智慧。Darpa 早就意识到了这个可被用在军事行动方面的机会。五年前，SRI 国际和领先的大学科研分包商达成了一个两千万美元的合同，项目是整合 AI 的子科目。目标是：一个移动，自动的战场指挥与控制助手。

在战斗中，不仅仅移动计算很重要。限制于他的尺寸，智能手机只有有限的信息捕捉和显示能力。用户必须从云获取数据，并且找到一个方法来在移动中作出决定。任何能够帮助用户快速化繁为简的 AI 应用都很必要。

获得了我的公司的资助，圣荷西（加州）启动 Siri，利用 Darpa 资助的 SRI 技术并开发出了一个允许用户通过网络语音接入服务的虚拟个人助手，比如餐馆预订和订购电影票。苹

果公司（AAPL）今年早些时候购买了这一技术（条件未公开）。现在可以期待很快能看见更多的基于 AI 的移动技术。不光是苹果，谷歌（GOOG）和微软（MSFT）已经把界面智能化作为开发重点。谷歌最近的关于 Android 上的 Voice Actions 的发布，被分析师认为是苹果购入 SIRI 的一个回应，或许被认为是即将到来的智能手机 AI 之争的第二轮交火。

那么 AI 的未来会如何？乐观主义者，比如撰写《奇异点已近》一文 的 Raymond Kurzweil，预见到了一个人机和谐的人工智能乌托邦。而悲观主义者，比如写出了预见性的《电脑天演论》（1863）的 Samuel Butler，预见着一个类似《终结者 3》中天网的反乌托邦的未来。但是现在，很明确的是，主流 AI 的时代正在到来。商人和投资者的或许更应该关心的是，“你的 AI 策略是什么？”

【原文链接】 【回到目录】

Ryan Blitstein：人工智能谱曲的新时代


Ryan Blitstein，芝加哥的自由撰稿人，以及媒体、政策研究中心 Miller-McCune 的特约编辑。

“

1987 年，当智能软件 Emmy 谱写的巴赫风格作品首演出时，伊利诺伊大学音乐厅的所有听众惊讶得目瞪口呆。两年以后，大多数人甚至无法分辨 Emmy 的“创作”和巴赫的原创。赢得赞誉的同时，Emmy 也引起了不安：如果机器编出的曲子可以与莫扎特的奏鸣曲媲美，那莫扎特还有什么独特之处？

”

从今往后，如果你经常听到音乐界传来有疑似莫扎特、巴赫的新作品问世，千万不要惊诧。这既不是古典音乐大师们的起死回生，也决非音乐界的“考古”新发现，这些作品出自音乐奇才 Emily Howell 的“手笔”。

Emmy 是 Experiments in Musical Intelligence （音乐智能实验）的简称（EMI），为了上口故称 Emmy。Emily Howell 是 Emmy 的升级版。这是美国加州大学圣克鲁斯分校的音乐教授兼作曲家大卫•科普开发的一款软件，被认为是目前最先进的人工智能音乐作曲系统。这款程序因为其具有以假乱真的能力而在音乐界引起了激烈的争论。现场聆听过科普先生表演的听众被这种新颖独特的作曲手法深深折服。科普先生为 Emmy 注入了生命活力，让它成了现代音乐界的弗兰肯斯坦 （科学怪人）。

致力于创造独树一帜、新潮有趣的当代古典音乐，是大卫•科普追求的梦想。

早在 1980 年，科普要为一部歌剧谱曲。很快预付的稿费被花完了，工作却没有多大进展。无论他如何费尽心机，合适的音符就是跳出不来。他甚至觉得自己已经丧失了审美判断力。在恐惧和绝望交加的时刻，他的目光转向了电脑。

科普先生在摆弄合成器制造各种音响效果的过程中，接触过谱曲软件。他受人工智能启发，觉得或许能找到一种方法，创作出有自己的风格的新作品。

他读遍圣克鲁斯图书馆人工智能的图书，又坐进课堂学习计算机编程。不久他开发出一套符合自己口味的简便作曲软件，但没过多久，他意识到这项工作难度太大。于是他把目标定得更加实际：从模仿巴赫风格的赞美诗做起。经过一年的努力，他编出的曲子已经达到



大学二年级生的水平。此后他又顺利引进和弦制作，不过乐曲的活力仍嫌不足，但对于软件作曲不啻是个小小的胜利，但作为制作原创性音乐的一种方法，还来路方长。

终于，功夫不负有心人。1987 年，当 Emmy 谱写的巴赫风格作品首次演出时，坐在伊利诺伊大学音乐厅的所有听众惊讶得目瞪口呆。两年以后，在圣克鲁斯巴洛克艺术节上，科普先生要求听众告诉他，哪首曲子是巴赫的原创，大多数人都无法给出正确的答案。

虽然 Emily 的作品给古典音乐学者留下了深刻印象，他们甚至无法相信这些曲子是出自电脑的手笔。科普先生因此而赢得不少音乐家和计算机专家的赞誉，但与此同时他的创作也引起令人不安的问题：如果机器编出的曲子可以与莫扎特的奏鸣曲媲美，那莫扎特还有什么独特之处？此外，人们还要问，这些大作到底有没有自己的灵魂，难道贝多芬等大师们的作品可以通过符号来操控？

一位音乐爱好者在不知道作品是由机器谱写之前，对它赞不绝口。半年后，他在加州大学圣克鲁斯分校出席科普先生的讲座，再次聆听了同一首作品，这次他告诉作者，音乐确实很美，但缺少“打动灵魂的深度。”

还有一位艺术爱好者霍夫斯塔特也有类似的看法，他认为音乐是人类传递深层情感信息的基本途径。而机器，无论拥有多么高超的运算能力，都无法拥有这种神奇的精神力量。

还有一种观点认为，科普的作品有抄袭之嫌。

科普对此是这样回答的：暂且把人类具有精神创造天赋的观点搁置一边，看看四周不计其数的音乐作品吧。“作曲家难道不是在听到些音符之后才创作新品的吗？那么第一个音符的作者是谁？是鸟，还是飞机？”

也许，确实有作曲家从鸟类汲取灵感，但科普认为，新作品肯定得到过其他作曲家的启迪，他们下意识地把听到的东西切成片段然后以新颖的方式重新拼接。不然怎么解释古典音乐的某种风格会传承三、四个世纪？为了证明自己的观点，科普对著名作曲家的作品进行逆向开发，寻找作品的修辞手法、语句和思想。

“没有谁的东西是完全原创的，在音乐创作上，我们听什么写什么。作曲家的任务就是审视历史聆听音乐，每个人都在向别人借鉴。不同之处在于你借鉴多少以及用多么高超的手段把这些东西进行组合。”科普认为。

科普先生是基于这样的理念来开发作曲软件的，他认为所有的音乐，或者任何具有开创性的工作，都是建立在已有知识的基础之上，或称为站在巨人的肩膀上（有人称之为剽窃）。其实我们创造的一切东西都是重新排列组合的产物。他的观点虽然有一定的道理，但他把人类的创造等同于简单的重新排列组合却让一些音乐家和艺术家大为不满，这不是在贬低自己的劳动吗？

在科普先生看来，即使是最受人尊敬的作曲家，在创造自己的东西时，或多或少受到过以往作品的启发。他正是由此豁然开朗，利用 Emily Howell 软件对著名作曲家的作品进行逆向开发，提炼作品中的构成要素。

实际上，科普先生的作品并非是建立在简单排列组合基础上的粗制滥造，而是人和机器的合作和交融。1981 年当他即将完成那部歌剧的时候，他已经为 Emmy 建立了一套数据库。科普先生是在用专业的元素和理念与机器进行音乐对话，机器在他的调教下，优化作曲手法，逐渐形成自己的风格，创作出属于科普先生的原创音乐。

虽然有音乐家和作曲家对他的工作持怀疑态度，但从现实的情况来看，科普先生确实开辟了机器谱曲的新天地，而且其作品一点也不亚于艺术大师。当他的作品登台亮相后，他得到了自己一生中最好的评价。90 年代后期科普教授名声鹊起，当时的《纽约时报》用许多版面报道了他的成果。

他还引起了对人工智能感兴趣的科学家们的关注。斯坦福大学人文学科电脑辅助研究中心的高级研究员埃莉诺认为科普的发现具有颠覆性意义。她认为，“科普先生揭示了许多有关音乐风格、作品定义和个人对音乐演化的作用等诸方面至关重要的要素，这些东西在他之前还没有被清晰地阐述过。他对于我们理解音乐确实作出了重要贡献，揭示了一些真正值得了解的东西。”

在与作曲软件相伴近 30 年之后，科普的作曲速度已经相当惊人。他刚才还躺在浴缸里构思，转眼之间他擦干身子，穿好衣服，来到电脑跟前，连续捣鼓 10 分钟，一首 100 多小节的乐曲便大功告成。不过他仍然有所迷惑不解，自己几十年的创新探索能否最终更接近于终极目标：谱写出流芳百世的不朽之作。

尽管引来不少争议，科普先生还是希望自己的探索能吸引更多作曲家的关注。他认为现在是百舸争流的时代，他相信自己的作为预示着新型音乐创作的未来：大批原创音乐电脑作曲家即将诞生。“如果能又快又好地写出听众喜爱的东西，就肯定能形成气候。”

“我相信这是迟早的事，我不认为作曲家都反应迟钝。他们最终会借助一切可用工具去追求音乐梦想，写出自己喜欢的音乐。或许开始会有所犹豫，但最终该发生的还是会发生，无论我们喜欢与否。”

目前，至少有一个著名的演唱组合（因为签署了保密协议，他不能说出其名称）请他用软件帮助他们谱写新作品。他指出这样的合作就像潘多拉盒子，一旦打开就会引起连锁反应。

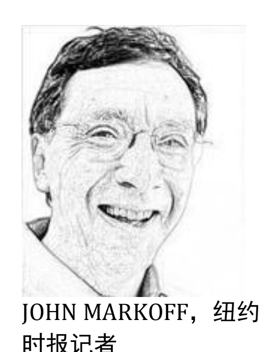
“我希望孩子们都能听到我的作品，演奏我的音乐，像我童年时期一样从音乐中获得乐趣，”他说。“如果这个愿望无法实现，那才是我完全的失败。”



【原文链接】 【回到目录】



John Markoff：让机器领会人类语言的“深度学习”


JOHN MARKOFF，纽约时报记者

> **现代人工神经网络由一系列软件组成，分为输入、隐藏层和输出几个部分。通过反复对图像或声音等规律模式进行识别，这些软件就可以得到“训练”。**

一些科技公司宣称，利用一项基于人脑识别规律模式理论的人工智能技术，它们在计算机视觉、语音识别，以及辨识可望用于制药的新分子等众多领域取得了惊人的成果。

这些成果在那些设计执行看、听、思考等人类活动的软件的研究者中激起了广泛热情。它们提供了科技前景，让人们有望制造出能够与人类交流、能够完成开车及工厂劳动等任务的机器，同时也让人们更加担心，能够取代人工的自动机器人即将问世。

目前，这种名为“深度学习”的技术已经被应用于以纽昂斯通讯公司（Nuance Communications）的语音识别技术为基础的虚拟个人助手“苹果语音助手”（Apple’s Siri），以及利用机器视觉来辨识地址的“谷歌街景（Google’s Street View）等服务软件。

但是，近几个月才有的新事物是深度学习程序不断提高的速度和精确度，这些程序通常被称作人工神经网络，或者简称为“神经网”，原因是它们与人脑的神经连结相似。

“深度学习的方法取得了一系列令人惊讶的新成果，”曾在贝尔实验室（Bell Laboratories）从事开创性笔迹识别研究的纽约大学（New York University）计算机科学家严恩•勒坤（Yann Legun）说。“这些系统在精确度上的巨大进步的确非常罕见。”

上个月，深度学习在中国天津的一次会议上得到了十分高调的展示。当微软（Microsoft）首席科学家理查德•F•拉希德（Richard F. Rashid）在巨大的礼堂里发表演说时，电脑程序对他的讲话内容进行了识别，还用英语把这些内容实时显示在了他上方的大屏幕上。

之后，他在讲完每句话之后稍作停顿，程序就把这些话翻译成了中文，同时还附上了模拟他嗓音的汉语配音，尽管拉希德从来都没说过汉语。这个展示震惊了观众，现场掌声雷动。

之所以能取得这个成果，部分是由于深度学习技术推动了语音识别精确度的提高。

负责监管微软在全球各地的研究机构的拉希德表示，虽然微软新语音识别软件的误差要比之前的版本低 30%，但“还是离完美很远”。

“现在的误差率是七分之一到八分之一，不再是四分之一到五分之一，”他在微软的官方网站上写道。但是，他表示这仍然是自 1979 年以来“精确度方面的最显著进步”，“而且，随着我们将更多的数据加入训练过程，我们相信自己能取得更好的结果。”

人工智能研究者非常清楚过分乐观的危险。长期以来，他们的研究领域一直充斥着不合时宜的爆发热情，随之而来的则是同样引人注目的倒退。

20 世纪 60 年代，有些计算机科学家相信，他们距离可行的人工智能系统只有十年之遥。而在 20 世纪 80 年代，一大批商业科技新兴公司纷纷倒闭，导致了一些人所说的“AI 之冬”（AI winter，即人工智能之冬——译注）。

然而，近期的成就给计算机领域的很多专家留下了深刻印象。举例来说，今年十月，在默克集团（Merck）赞助的用于寻找有可能衍生新药的分子的软件设计竞赛中，一个与加拿大多伦多大学（University of Toronto）的计算机科学家杰弗里•E•欣顿（Geoffrey E. Hinton）一起从事研究工作的研究生小组获得了头奖。

利用深度学习软件，他们从描述了 15 种不同分子的化学结构的数据组中挑出了最可能成为有效药物助剂的那种分子。

这个成果尤其令人震惊，因为这个团队是事到临头才决定参赛的，而且他们设计软件的时候，对分子和目标之间的联系并没有特别深刻的了解。同时，这些学生面对的是一个相对较小的数据组，而神经网络通常要在非常大的数据组中才会有良好表现。

“结果真的非常惊人，因为这是深度学习方法首次胜出，更值得一提的是，人们根本想不到它会在这样一个数据组中取胜，”预测分析公司 Kaggle 的首席执行官及创始人安东尼•戈德布卢姆（Anthony Goldbloom）说。该公司经常组织数据科学竞赛，包括由默克赞助的上述竞赛。

规律模式辨识领域所取得的成就不仅将对药品研发产生影响，还将对市场营销和执法等诸多方面产生影响。例如，随着精确度的提高，市场营销人员可以通过梳理关于消费者行为的大型数据库来获得更准确的消费习惯信息。面部识别技术的进步也会降低监察技术的成本，使之更加普及。

人工神经网络的理念源于 20 世纪 50 年代，旨在模拟人脑吸收信息并从中学习的方式。近几十年，64 岁的欣顿博士（19 世纪数学家乔治•布尔 George Boole 的玄孙，布尔在逻辑领



域的工作构成了现代数码计算机的基础）率先推出了一些强大的新技术，用来帮助人工神经网络识别规律模式。

现代人工神经网络由一系列软件组成，分为输入、隐藏层和输出几个部分。通过反复对图像或声音等规律模式进行识别，这些软件就可以得到“训练”。

在现代计算机日益增长的计算速度和计算能力的帮助下，这些技术推动了语音识别、新药研制和计算机视觉等领域的快速发展。

最近，在一些特定的有限认知测试中，深度学习系统的表现甚至超过了人类。

例如，卢加诺大学（University of Lugano）瑞士 AI 实验室（Swiss A. I. Lab）的科学家开发的一个程序在去年的一场规律模式辨识竞赛中胜出，在德国交通标志的数据库中分辨图像时，其表现超过了参与竞赛的其他软件系统和人类专家。

在包含 5 万张图像的数据组中，获胜的程序精确地辨识出了其中 99.46%的图像；而由 32 人组成的人类参赛小组所取得的最好成绩是 99.22%，人类的平均水平则是 98.84%。

今年夏季，谷歌的技术人员杰夫•迪安（Jeff Dean）和斯坦福大学（Stanford University）的计算机科学家安德鲁•吴（Andrew Y. Ng）把 1.6 万台电脑连在一起，使其能够自我训练，对 2 万个不同物体的 1400 万张图片进行辨识。尽管准确率较低，只有 15.8%，但该系统的表现比之前最先进的系统都要好 70%。

欣顿博士领导的研究项目最突出的方面之一是，研究工作基本不受专利权的限制，也没有因为争夺知识产权而发生激烈的内部斗争，虽然这种斗争在高科技领域很常见。

“我们很早就决定不利用这个挣钱，只是想把它推广开来，影响到每个人，”他说。“这些公司都对这一点感到非常高兴。”

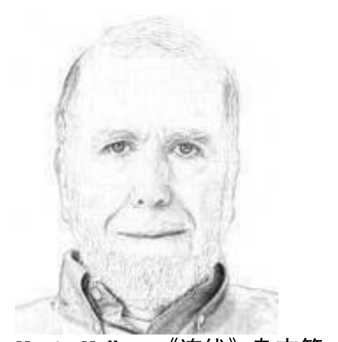
更强的计算能力，尤其是图形处理器的兴起，带来了深度学习的快速发展，说到这一点时，他补充道：

“这项技术的妙处在于，它的发展模式非常漂亮。你基本只需要不断扩大它的规模、加快它的速度，它就会越变越好。现在它是不会走回头路的。”

【原文链接】 【回到目录】



Kevin Kelly：比人类更强：为什么机器人终将也必将接手我们的工作？



Kevin Kelly，《连线》杂志第一任主编，作家、摄影师

> “
> **我们不是要与机器为敌。如果我们跟机器比，我们会输。我们要与机器并肩作战。**
> ”

试想一下，如果明天有 70%的美国人被解雇，他们会去干什么呢？如果真像这样有一半以上的劳动力被炒鱿鱼，这个经济体恐怕也会不复存在吧？

但是 19 世纪初工业革命时期，劳工们就曾像这样被慢慢解雇。200 年前，70%的美国劳力在农场工作。现在，自动化技术抢走了 99%的工作，机器取代了他们（和他们的农用畜力）。但是，被解职的劳力并没有闲下来。自动化技术反而在新的领域创造出上亿的工作岗位。原来的农民加入了工人大军，在工厂生产农具、汽车和其他工业产品。从那时起，一波一波的新岗位相继涌现——设备维修、胶印、食品化学、摄影、网页设计——每一项都植根于先前的自动化技术。今天我们大部分人的工作都是 19 世纪农民根本无法想象的。

也许我们很难想象到了本世纪末，今天 70%的工作可能会被自动化技术取代。是的，亲爱的读者朋友，你们的饭碗也会被机器抢走。换句话说，被机器人取而代之只是时间问题罢了。

革命已经开始

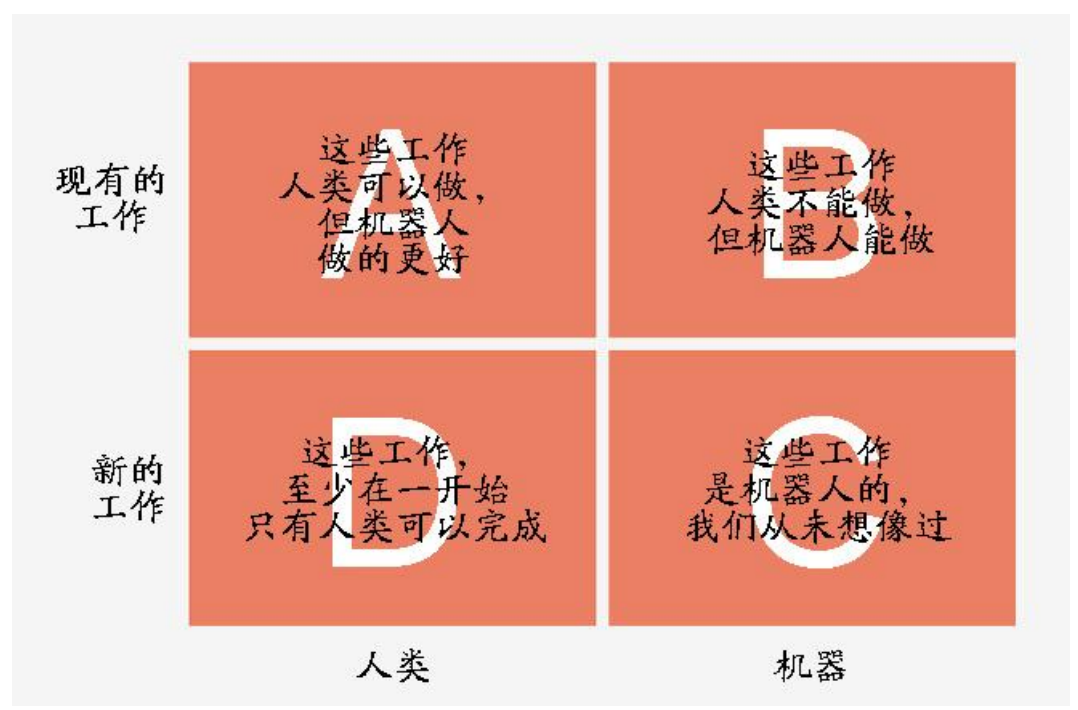
这场巨变是由第二波自动化革命引领的，其核心是廉价的传感器和人工识别、机器学习、分布式智能技术。这种深层次的自动化，将触及从体力劳动到脑力劳动的所有岗位。

首先，对已经实现自动化的工业部门，机器会进一步巩固它们的地位。机器人会取代组装流水线上的工人，然后再取代仓库里的工人。它们可以搬运箱子，对箱子进行分类，然后装上卡车。机器人还将继续进军蔬菜水果采摘业，直到把人类全部挤出普通农场。药店后台有一位发药机器人就够了，药剂师只要关注病人的咨询就好。接下来，打扫办公室和学校这种需要多些技巧的工作也会被夜班机器人抢走，它们可以从打扫门窗干起，最终学会打扫厕所。高速长途货运则会由固定在驾驶室的机器人完成。

机器人还将继续进军白领工作。我们的很多机器上都已经有了人工智能，只不过我们没有意识到。以美国叙述科学公司的一款软件为例：这种软件可以根据比赛情况撰写体育赛事新闻，或者根据网络上点点滴滴的文字通报公司股票每天的表现。任何文书工作都可以由机器人胜任，包括医学领域的许多工作。即便在那些不能简单用文书概括的医学行当，比如外科手术，机器人的身影也越来越多。任何有关海量信息处理的死板日常工作都可以自动化。无论你是医生、律师，还是建筑师、记者，甚至是程序员也一样：机器人革命是大势所趋。

人机关系的四个分区

为了便于理解机器人是如何拿走我们饭碗的，现在我们把人机关系分为四类，如右图所示：



上下两行分别代表我们现有的工作或新工作，这些工作将由机器人接手，左右两列分别代表(最初看来)属于人的工作或属于机器的工作。

首先来看区域 A：这些工作人类可以做，但机器人做得更好。如果人类愿意花费大量劳力，是可以纺出棉布的；但是自动织布机可以织出更完美、更便宜的布。现在想要去买手工布的唯一理由就是你想要保留一些人为的缺陷。但是当你开车开到每小时 70 英里的时候，你不会希望你的车犯毛病——所以在造车过程中，我们觉得越少人碰我们的车越好。

但是事情并不是这么简单，我们总觉得计算机和机器人不太可靠。自动驾驶仪就是一种电子脑，它可以独立驾驶 787 喷气式飞机，但我们还是一反正常逻辑，把人类飞行员放在驾



驶舱里监督自动驾驶仪，为了“以防万一”。20 世纪 90 年代，电脑估价已经全面取代了人工评估抵押贷款，许多报税工作都已经由电脑来完成。还有日常 X 光检查和审判前的证据收集。以上这些工作，以前都是由领高薪的聪明人来干。对于机器人生产产品，我们已经完全认可了，不久我们将会接受机器人的智能和服务。

再来看区域 B：这些工作人类不能做，但机器人能做。一个简单的例子：没有任何辅助，人类很难做出一枚黄铜螺丝钉，但是自动化技术可以在一个小时内生产出 1000 个一模一样的螺丝钉。没有自动化技术，我们不可能做出电脑芯片。做一块电脑芯片需要一定的精度、控制和极高的注意力，而我们的肉体不具备这些条件。同样的，无论教育程度如何，没有人可以迅速搜索世界上的所有网页，从中找出自己想要的一页，比如找出昨天加德满都的鸡蛋价格，就算是一群人来搜索也不行。每当你点击搜索按钮的时候，你都是在指挥机器人干活，做一些我们人类没法独立完成的工作。

尽管对机器人抢走人类饭碗的抱怨占据了各大头条，但是机器人和自动化也带给我们极大的好处：一些我们不能做的工作需要由它们来完成。我们没有计算机辅助检测（CAT）设备那种注意力，不可能一个平方毫米一个平方毫米地扫描搜索癌细胞。我们没有毫秒级的反射速度来把融化的玻璃吹成瓶状。我们也没有万无一失的记忆力，不可能记住美国职业棒球联赛上的每一个球，更别说还要立刻计算出下一个好球的概率。

我们并没有把“好工作”交给机器人。多数情况下，我们把自己不干的活儿交给机器人。这些活儿如果机器人不做就没有人做了。

现在我们来看区域 C：这些新工作是自动化的产物，其中一些工作我们自己都不曾想到。机器人代替人类的最伟大之处就在于：在机器人和计算智能的帮助下，我们能够做一些 150 年前从未想象过的工作。我们能穿过肚脐切除肠道肿瘤，能对我们的婚礼进行录像，能在火星上开车，也能让远方的朋友把图案隔空发送过来，然后在纺织品上打印出来。我们现在所做的上百万种新鲜事可能会把 1850 年前的农民吓得眼花缭乱，这些事情有时还能换来报酬。这种转变不仅仅是一个从难到易的过程。这些梦想的实现主要依赖于完成这些任务的机器，依赖于它们本身的能力，是机器给我们创造了工作机会。

需要特别指出的一点是，自动化技术创造出来的大部分新工作只能由其他自动化技术来完成。当我们有了谷歌这样的搜索引擎，我们又开始交给它上千条新任务。谷歌，你能告诉我我的手机在哪儿吗？ 谷歌，你能领抑郁症病人去找卖药的医生吗？谷歌，你能预测下一次病毒性传染病什么时候爆发吗？技术进步是一视同仁的，它给人类和机器都创造了机遇与选择。



可以肯定，2050 年最赚钱的行业是建立在一些现在尚未发明的自动化技术和机器上的。也就是说，我们现在无法预见这些行业，因为催生这些行业的机器和技术还尚未出现。机器人会创造出我们想象不到的工作。

最后，我们来看区域 D：这些工作在一开始只有人类可以完成。一件人类可以做到但机器人（至少在很长一段时间内）做不到的事情就是，猜猜人类到底要做什么。这不是一个文字游戏；我们的欲求来源于我们先前的发明，这样说起来，判断人类想要做什么就是一个循环问题。

当机器人和自动化技术接手我们大多数的基础工作以后，我们的粮食，衣服和居住问题就可以比较轻松地解决。然后，我们才有时间来思考，“人类是什么？”工业化不仅延长了人类的平均寿命，它还让许多人意识到，人类可以是芭蕾舞演员、专业音乐家、数学家、运动员、时尚设计师、瑜伽大师、同人小说家，或者是其他名片上标注的职称。在机器的帮助下，我们胜任了这些角色；但显然，随着时光的流逝，机器又会把这些工作做得更好。接着，我们又重新出发，再次思索“人类应该做什么？”这个问题，找出更多的答案；而机器人又要经过几代才能回答这一问题。

尽管大部分工作都是由机器人完成的，但后工业时代的经济会继续扩张，因为我们未来的工作中，有一部分就是寻找、创造和完成新的事情；当然，这些事情随后又会成为机器人手中的重复性工作。

每个人都有机会得到自己的机器人，但仅仅拥有机器人是不能保证成功的。只有那些在管理优化机器人上开拓，在训练机器人完成任务上创新的人才会成功。制造工厂的地域集中也是取得成功的关键因素，不是出于劳动力成本的考虑，而是因为人们的专业技能因地域而有所不同。人与机器人形成一种共生关系。人类的任务就是不断给机器人分派工作——这是一项永无止境的工作。所以最起码，我们总有一项工作可做。

要与机器并肩作战

我们不是要与机器为敌。如果我们跟机器比，我们会输。我们要与机器并肩作战。

将来你的薪酬高低取决于你跟机器人协作的好坏程度。90%的同事会是从未见过的机器，而没有它们，大多数事情你都做不了。你做的事和它们做的事不再是截然不同的。你不会觉得自己做的事情是一份工作，至少一开始不会，因为所有无聊的事情都是机器人做的。

我们应该让机器人接手。它们可以去做我们正在做的事，做得比我们更好。它们可以去做我们做不了的事。它们可以去做我们从未想过要去做的事。同时，机器人可以帮我们给自己找到新工作，扩展我们存在的价值。机器人让我们有精力专注于怎样活得更像个人。



【原文链接】 【回到目录】



【原理·人工智能】

韩彦文：超越逻辑


韩彦文，科普作家

> **“事情似乎是这样的：计算机试图在用一种勤能补拙的方式与人类抗衡，通过不厌其烦地将最简单的逻辑重复重复再重复，来完成人类几乎一蹴而就的分析过程。”**

一个人和一台机器分别呆在两间封闭的屋子里，从外面看不见也听不见什么，另一个人用传真的方式对他和它进行提问，一段时间后，如果提问者没有分辨出哪一个是人哪一个是机器的话，我们就可以认为这台机器能像人一样思考。

这是英国天才数学家阿兰•图灵在 1950 年提出的判断计算机是否具有智能的测试。他预言，总有一天计算机可以通过编程获得与人类竞争的智力。1997 年，国际象棋棋王卡斯帕罗夫输给了超级计算机“深蓝”，没有体力消耗、注意力永远集中、不受任何心理影响，面对这样一个敌手，卡斯帕罗夫心力憔悴。在人类引以为自豪的智力游戏上击败人类，图灵的预言实现了。可难道我们必须要承认这台冷冰冰、硬邦邦的机器是智能的，而且比我们还要聪明？

智力之谜

至少国际象棋是下不赢它了，但可以肯定，面对我那脑筋早已不灵光的老祖母都能听懂的笑话，它会令人扫兴地无动于衷，向它投以爱意的目光也得不到任何的回应。那么凭什么说它是智能的？智能、智力是指什么？当我们说一个人比另一个人更聪明时是什么意思？智力可以量化吗？如果可以的话，是否能够说一只猫比一只蝴蝶更聪明，一只蝴蝶比一只蚯蚓更聪明？如此一来，某种生物应该拥有最少量的智力，这点智力能供它做些什么？而认得镜中的自己，并自信地问出：“魔镜魔镜告诉我，我是不是世界上最美的人？”又需要多少份的智力？

这些问题令人着迷，不仅因为有趣，还在于猫拥有爬树的能力，但不会思考自己为什么能够爬树，蝴蝶拥有飞翔的能力，但不会思考自己为什么能够飞起来，我们蒙造物之青睐，

进化出思考的能力，于是会去思考猫如何爬树，蝴蝶如何飞翔，最终还要不可避免地来思考思考本身：智力究竟是什么？它是如何产生的？

在西方，柏拉图首先认识到智力是由大脑产生的。2000 多年后的 1904 年，一系列认知实验表明，同一个人在不同的认知任务中都会表现的很出色，英国心理学家斯皮尔曼由此提出在我们的智力活动中，有一个通用因素（general factor）发挥着决定性作用。在将这个通用因素量化后，我们得到了评判一个人聪明程度的标准——智商（IQ：intelligence quotient）。

虽然以前也可以通过经验对一个人的智力水平做出八九不离十的推断，但现在我们有了更简单的办法——给他做一套智商测试题，看看分数如何。仅凭一个分数就对一个人聪明与否下定论似乎显得过于草率，但事实是智商已成为这个世界上继阶级、种族之后又一个将人与人区分出差别的工具，而且不得不承认它很有效。

正如在无数个苹果掉落于地上之后牛顿才发现万有引力一样，从经验上升到理论需要一个漫长的过程。在智商这个评判智力水平的标准大行其道之时，我们对智力是什么以及它是如何产生的却仍然一无所知。当然，这些问题并没有被科学家遗忘，只要时机成熟，他们就要有所作为。

2000 年，脑科学家约翰•邓肯在用不同类型、不同难度的智商测试题为难受试者时发现，大脑为了完成任务，总是会征召同一块皮层区域——侧额叶（lateral frontal cortex）。这似乎表明侧额叶的功能与所谓的通用因素密切相关，换句话说，侧额叶在智力产生过程中起到了关键性作用。从整个大脑到一块皮层，智力的来源范围缩小了，但随后的研究又将这个范围扩大了一些，执行任务者变为一个额叶——顶叶网络，其成员包括侧前额叶（lateral prefrontal cortex）、前扣带回（anterior cingulate cortex）和后顶叶（posterior parietal cortices）。后顶叶取代侧额叶成为关键角色，驱动整个网络的运转，并且与智商高低直接相关。天才不是像我们以为的那样，在酝酿奇思妙想时调动了普通人未曾开发过的神秘脑区，而只是更充分地利用了后顶叶来解决问题。（负责智力活动的额叶——顶叶网络（红色））

不过智商高就一定是天才吗？天才就一定智商高？智商只有 75 的阿甘（电影《阿甘正传》的主人公）用他奇迹般的一生质疑了这一点。当然这只不过是个电影，而现实生活中“雨人”的事迹是无可置疑的，智商一般在 35 至 70 之间的他们往往无法握住碗筷吃饭，穿衣叠被困难重重，但却可以在 30 秒内准确算出 2 的 64 次方是多少；在初次听到柴可夫斯基第一号钢琴协奏曲几小时后，将其行云流水般毫厘不差地弹奏出来……如果说这是一种异常情况，缺少普遍性的话，正常人中智商不高却有所成就的人物会更具说服力，比如现任美国总统。



看来智商并不代表一切，平日里不大会有人一直在推断一组数字的末尾应该是多少，或是根据前面的图形去决定最后一个的样子，而斯皮尔曼的认知任务和智商测试题却只考虑了这一种能力——分析、逻辑、推理。但事实上，我们并不必总是深思熟虑、条理清楚，瞬间的灵光闪现和大而化之的宽容、幽默很多时候会更有效，而这些也都是人类伟大的进化成果，需要被扩展到智力的概念里。

应运而生的是斯腾伯格的三元智力理论，他将智商所代表的通用因素归属为分析智力，除此之外还有创造性智力和实践智力。前者包含了灵感、直觉、想象力等，在诸如即兴作诗、给一幅卡通画加上标题时要用到。创造能力主要跟右半球有关，这半个大脑负责音乐、绘画、空间几何以及想象和综合等功能，一部分“雨人”的天分就是在左脑受损后才意外获得的。后者指的是解决实际问题和做出决定的能力，因此当我们嘲笑小布什的低智商时，不妨考虑下他是否在这方面有过人之处。

从柏拉图时代到现在，人类的进步仅限于把智力的诞生地从笼统的大脑锁定至特定的皮层区域，至于大脑是如何通过智力活动为爱因斯坦、达尔文、秦始皇、毛泽东带来巨大成就的，我们茫然无知。

全文阅读请点击链接。

【原文链接】　　【回到目录】



黄永明：模拟人脑，迄今最大规模

黄永明，南方周末记者

一个包含 5300 亿个神经元和 100 万亿个突触的人造“大脑”诞生了。但它仍然不能思考或是感受。“这有点像是拥有了世界上最大数量的乐高块，却对于能从中得到什么毫无头绪。真正的艺术不在于买乐高块，而在于知道如何把它们组装起来。”《纽约客》评论说。

2012 年 11 月 14 日，在美国犹他州盐湖城的 2012 超级计算大会上，IBM 的计算机专家提交了一份标题是“1014”的报告。报告所描述的研究被媒体称为“认知计算的里程碑”。使用了世界上运算速度最快的 96 台计算机，研究人员制造出了包含 5300 亿个神经元和 100 万亿个突触的人造“大脑”。

这是迄今为止对大脑的最大规模的模拟。它的初始目标是模拟出实验室中常见的恒河猴的大脑。有网友戏称“世界上第一台人工智能计算机诞生后的第一件事会是要一根香蕉吃”，但现在这个“大脑”只是“神经形态工程学”意义上的。

这里所说的神经元和突触并不是生物学意义上的。对于计算机来说，神经元就是计算中心，每一个神经元能够从周边成千上万的其他神经元那里接收到输入信号，对数据进行分析处理，然后再发射出去；突触是用于连接神经元的，它是记忆和学习发生的地方。

换句话说，神经形态工程学是受生物脑的功能和结构的启发，来改进计算机。比如说，传统的计算机依靠数量巨大的晶体管的打开与关闭来运行，要求接近完美的动作。这意味着电力的消耗。人脑只需要 20 瓦的电力，但是如果用传统方式建立与人脑规模相当的计算机，其用电量将相当于一个小城的用电量。之所以会这样，一个原因就是人脑是“事件驱动”的，只有当某些神经元被激活时才会较为消耗能量。而且，人脑突触的失误率非常高，在 30%到 90%的情况下会动作失败，但人脑工作得很好。假如是传统的计算机，晶体管一次失误可能就会导致死机。

全文阅读请点击链接。

【原文链接】　　【回到目录】



经济学人：人工智能，模仿昆虫智慧

“

大脑的工作原理可能更像神经网络，而不是中央集权式的发号施令。甚至那些复杂的认知功能，如抽象理论与感觉，也是那些神经元在互动，就像是在跳“摇摆舞”而已。这或许正是人工智能的核心。

”

科幻小说家用来制造恐怖气氛的常见怪物之一就是经过进化，拥有高级智能的蚂蚁变成了地球的主宰。对蚂蚁及其他一些群体性的昆虫而言，它们的有目的的群体行动看起来似乎很有智能。当然，这可能只是一种假象而已。但是对于计算机科学家而言，这已经足够进行研究了。目前，寻求通过研究人的大脑获得人工智能灵感的方向已经失败。与之对应的是，通过研究蚂蚁的行为也许可以了解真实世界运行原理的关键。

1990 年代，蚂蚁首次得到了软件工程师的关注。一只蚂蚁看来是无所作为的，但是当它们聚焦成群后，能力就得到了极大的提升。他们可以极为复杂的工作，如筑起巢穴，维持其正常动作，在里面填满食物等。这个事例提醒了 Marco Dorigo，他是一名布鲁塞尔大学的研究员，同时他也是昆虫群智能研究领域的创始人。

Marco 特别对于蚂蚁如何得到食物与巢穴之间的最短路径感兴趣。这也给了解决计算机界一个经典难题—推销员路径问题的提示。这个难题的简单描述是：给定一个城市列表，同时给定这些城市之间的距离，要求得出推销员访问所有城市的最短路径。随着城市数目的增长，这个问题变得极为复杂。计算机解决这个问题，需要极长的时间，以及极其强大的计算能力。当然研究这个问题可不仅是为了替那个推销员规划旅行路径，而是因为这个问题在许多领域都有其重要性，如设计集成电路、聚合 DNA 分子等，这些问题简化归约成了推销员路径问题。

蚂蚁用被称为信息素的一种化学信号得到了这个问题的一个解决方案。当一个蚂蚁找到食物后，它会将其带回巢穴，并且在沿留下信息素指引同伴。蚂蚁路过的越多，则该路上留下的信息素也就越强烈。信息素挥发地很快，一般来说，当食物被搬进巢穴后，信息素所留下的痕迹就挥发殆尽了。快速挥发的信息素意味着，对于蚂蚁而言较长的道路没有短一

些的路有吸引力。所以，信息素极大地放大了蚂蚁个体的智慧，将整体智慧提升到一个更高的级别。

集群智慧

1992 年，Dorigo 与他的团队开始研究 ACO（Ant Colony Optimisation 蚂蚁群体行为优化），这个算法模拟了一群蚂蚁在田野上漫步并且留下信息素的行为，寻求解决“推销员路径”问题的答案。ACO 系统很好地解决了“推销员路径”问题。从那时起，这个算法不断成长，成为一个系统性的算法，被应用在许多特定的问题之中。

取得了巨大的成功。 Migros，是一个瑞典的超市供应链主管，Barilla，是意大利通心粉供应商，他们日常工作就是将货物从集散中心发到各个零售店，他们的发货使用的软件是 Ant Route（蚂蚁搬家）。这只是 Ant Optima 公司开发出的其中一个软件。而该公司也是 IDSIA（卢加诺人工智能协会）的一个子公司。IDSIA 是欧洲领先的昆虫智能的研究中心。每天早上，“蚂蚁”根据货物的质量，目的地， 发送窗口，以及可用的卡车等，计算出货物发送队列的最优路径。根据 IDSIA 与 AntOptima 的主管 Luca Gambardella 的计算，每天有 1200 辆卡车被发往各地。尽管计划每天都在变动，但每天计算的整个过程不会超过 15 分钟。

类蚂蚁的算法也被广泛应用在通信网络的路由计算问题之中。IDSIA 的另外两个研究员，Dorigo 与 Gianni Dicaro，开发出了蚂蚁网络，原理很类似于蚂蚁寻找食物的过程。这是一个路由协议，信息包从节点之间短转移，转移过程中，会留下标注着“质量”的信号。其他的信息包会沿着前面的包所留下的包前进。在小范围的网络测试表明，“蚂蚁网络”的表现要比现在的路由协议都要好。它有更好的适应能力，可以应付复杂的状况（如突发的流量），对于节点错误也有更好的适应性。据 Di Caro 透露，有许多大公司对“蚂蚁网络”很感兴趣，但是，如果要使用“蚂蚁网络”，则现存的网络设备需要更换，这是一笔极大的投资。但蚂蚁网络可能被用在许多场合，如移动基站，军事通信以及其他一些场所。

路由，包括负载和路径，是数学上称为离散变量的问题，研究对象是一个衡定的大值。而对于连续变量（相对于离散）的难题，特别是那些解决方案有无穷多的问题（如寻找机翼形状），其他的昆虫智能也能工作得很好。特定的昆虫优化（PSO），如 James Kennedy 与 Rusell Eberhart 在 1990 年代中期发明的算法，对鸟类更感兴趣。当你在阳台上放置一个饲鸟器的时候，过很久才会有第一个来进食的鸟，但是紧跟着，其他的鸟很快翼会飞来。PSO 试图重建这个过程。人工鸟开始是随意地飞行，但是随时关注着其他的鸟，它一般是紧跟着食物的鸟。现在 PSO 有 650 个测试应用程序，涵盖了图像、视频分析、天线设计，还包括医学诊断系统，以及工业设备中的错误检测系统等。



数字蚂蚁和鸟，在许多问题上工作地很好，但是 Dorigo 博士现在着力罢考一个新的课题：机器人。 一些小而便宜的机器人可以通过协作完成那些大且昂贵的机器人同样的工作，而且更为灵活与健壮，意思是如果一个机器人失效，昆虫机器人还是可以正常工作。 今夏，他将演示他的“Swarmandoid”项目。这个项目基于三种小而简单的机器人，每种都有不同的功能，通过协作来探索周围的环境。眼睛机器人负责观察并定位感兴趣的物体。 然后行走机器人会载着触手机器人到眼睛机器人所标识的地方。触手机器人会捡起感兴趣的物体。 然后他们就返回起始地点。

所有的这些工作的完成都没有预先制定的计划，当然也没有中央指挥机构来协调。 所有的这些都基于单个机器人之间的互动之上。据 Dorigo 博士透露，这种昆虫机器人适用于生存救援，例如，定位幸存者、在火灾中找回有价值的商品。

智能

Swarmanoid 机器人的外形与激发创造他们的灵感的原型（鸟）相去甚远，但是昆虫给程序员很多灵感。Dorigo 博士的团队现在致力于研究一个系统，这个系统允许机器人检测团队的一个成员何时发生故障。这个设计的灵感来自于萤火虫同步它们的闪光，可以使整棵树的萤火虫闪亮或熄灭。机器人也做同样的事，如果一个灯光熄灭，由于同步，其他的机器人很快会有所反应，而且可以孤立不与系统同步的机器人，以避免其对系统带来麻烦，或者可以呼叫基地来强制执行。

所有这些都很鼓舞人心。但是，任何对人工智能的难题感兴趣的人都会不由自主地想到人的思维，会惊叹于人的大脑的能力。不过，大脑的智慧绝不是人工智能所无法企及的。Dorigoand 博士与他的同事已经在研究大脑智能的征途中，他们找到了当前研究水平下对于大脑思维的一个卓越模仿对象。

例如，根据认知科学与技术协会的 Vito Trianni 的研究，罗马的蜜蜂选择回巢路径的方法看起来就像是他们存在着智慧一般。侦察蜂在适当的地点探索一个区域。 当他们发现一个好的地点时，他们返回巢穴开始跳摇摆舞（用来显示发现了花蜜丰富的花源），其他侦察蜂也会蜂拥而至。花源地的花蜜质量越高，摇摆舞的频率和幅度就越大。直到更多的侦察蜂赶过来，带来足够多的工蜂为至。 用神经网络来代表蜜蜂，用电子活动来代表摇摆舞，现在对于刺激及反应就有一个良好的描述了。

昆虫认知理论的提出者，如 Trianni 博士，认为大脑的工作原理可能更像神经网络，而不是中央集权式的发号施令。他们认为，甚至那些复杂的认知功能，如抽象理论与感觉，也是那些神经元在互动，就像是在跳“摇摆舞”而已。他们谈论的是蜜蜂的智能，但是，他们可能没意识到，他们以一种隐喻的方式触及了人工智能的核心。




我在中國
論壇
Co-China

John Searle：《图灵测试：五十五年之后》（《新世纪的哲学》第三章节选）


John Searle 约翰·罗杰斯·希尔勒，加州大学伯克利分校哲学教授。

> **“人工”一词乃是系统性地混乱的，因为“一个人工的X”既可以是一个实在的、依靠技术生产出来的 X，也可以是一个并非是实在的 X，而只是对 X 的一个仿造品的东西。**

一、图灵测试的不同解释方法

尽管图灵最初的文章（Turing, 1950）是以清晰简洁的散文写成的，但其提出的论断却可以用不同的方式来解释。在本文中，我并不是要探讨图灵真实的意向是什么，而是聚焦于图灵测试之结果的三种不同解释方法，它们在图灵测试的应用当中都很重要。为了本文考虑，我将假设这一测试本身是清楚明白的。我的探讨将考虑下述问题：我们如何解释一个明确的结果？根据一种自然的解释，该测试使我们能够说明我们是否已经成功地模拟了人类的认知能力，即表现为思考的某些形式的人类智能行为。如果机器能够这样来执行任务，使得一个专家无法分清楚究竟是机器在执行任务还是一个足以胜任的人类在执行任务，那么此机器就成功地模拟了人类的智能行为。事实上，如果在人工智能（Artificial Intelligence，AI）方面我们的目标是制造出一台可以成功地模拟人类智能的机器，那么图灵测试就给了我们判断成功与失败的标准。我看不出人们怎么会反对这样的一个测试。如果问题是我们实际上是否模拟了（也就是仿造了）人类行为，那么平凡地来说，如此解释的图灵测试看起来是正确的：如果你无法分别原型和仿造品，那么后者就是一个成功的仿造品。

但还有另一个解释图灵测试的方法，它所给出的结果在我看来并不令人满意，并且事实上乃是错误的。图灵测试时常被解释为是哲学或逻辑行为主义的一个应用。事实上这或许是正确的，因为在图灵写作的年代，行为主义乃是心理学和哲学中的主流理论。习惯上逻辑（或哲学）行为主义和方法论行为主义会有所区别。逻辑（或哲学）行为主义乃是哲学中的一个论题，它认为心智现象乃是由行为和行为倾向（disposition）所构成的；方法论行为主义乃是心理学中的一个研究方案，它主张研究心理学的适当方法乃是研究人类行为。



当我说“行为主义”时，我所说的乃是逻辑行为主义。根据行为主义概念，人类的心智现象只不过是由行为和行为倾向所构成的，也就是说可以被还原为行为和行为倾向。对于某些认知现象来说，如果一台机器能够产生和人类完全相同的行为，那么此机器就完全像人类那样具有此认知现象，这乃是行为主义的严格的逻辑推论。基于第二个解释，我们应当把图灵测试看作是行为主义原理在人工智能方面的应用。如此说来，该测试表明，难以区别机器行为和人类认知行为这一点并不只是机器之成功仿造或模拟的证明，更是在机器中适当的认知之存在的决定性证明。怎么会这样呢？就其原始形式而言，行为主义乃是这样一种观点，即外在行为的表现不只是系统内在认知之存在的证据，相反，行为和在适当的环境下表现出该行为的这一倾向就构成了心智。凡涉及心灵的地方，就只有行为和行为倾向。

当在哲学探讨中语言模式最为强势时，行为主义十分繁荣，并且行为主义典型地被陈述为一个关于心理属性的论题。行为主义者都典型地认为关于心智状态的陈述都可以完全翻译为关于行为的绝对性和假设性陈述。引入假设性陈述概念是为了解释倾向概念。举个例子，说一个人具有一个特定的信念，就是说他/她现在以一个特定的方式行动（绝对性的），或者如果如此这般的条件出现，他/她将以一个特定的方式行动（假设性的，倾向性的）。基于这一解释，图灵测试乃是行为主义的一个推论，因而如果图灵测试是错误的，那么行为主义就是错误的。要区别图灵测试的这两个解释，我将使用数年前我在区别两种人工智能时所引入的一个用辞。我将称第一个图灵测试为弱图灵测试，根据这一图灵测试，通过测试就是成功仿造或模拟的决定性证明。我将称第二个图灵测试，或者说行为主义图灵测试为强图灵测试，根据这一图灵测试，通过测试就是心理现象之存在的决定性证明，因为构成通过测试这回事的行为构就成了心理现象。

很明显行为主义或多或少是错误的，如果你稍微反思一下你自己的下述心智过程就能够看到这一点：想一下你将在那里度过暑假的地方，或者感受一阵疼痛。具有心智状态是一回事；在你的行为中展现这一心智状态是另一回事。具有心智状态而不在行为中展现它是可能的，同样具有一个行为而没有相应的心智状态也是可能的。我发现很难想象人们会无法看到这些明显的要点。但是，数个知名哲学家确实否认这些要点，并且断言心智生活可以还原为行为和行为倾向。

行为主义者时常碰到的一个困难是解释这样一个明显的事实，即在涉及智能行为时总是典型地有两件事发生，就是内在的心智过程和这些心智过程在行为中的表达。比如现在我具有思考艾伦•图灵的著名论文这一心智过程，而这些思维过程在我的行为中表现为在一台计算机上键入我的想法。纯粹的行为主义者英勇地否认在这一情况中有两件事发生，即内在的心智过程和外在的行为。只有智能行为，此外无他。如果你看一下图灵的文章，尽管看起来他认为外在行为是对其他事物的决定性测试，但一些认知过程却是在内部发生的。但



如果那就是解释该测试的正确方式，那么它就总是屈服于这样一个异议，即相同的外在行为或许是由两个截然不同的内在因果机制所引起的。我们可以制造一台能够复制人脑的外在行为输出而没有任何内在思维过程的机器。（比较：我们可以制造一台和汽油发动机具有相同功率输出的电动发动机，但这无法证明此电动发动机具有内部燃烧。）所以现在我们有了三种不同的方式，来解释图灵测试的积极结果。

（1）弱图灵测试：我们并不在乎系统内部所发生的事情。我们只是想复制智能行为。如果机器通过了图灵测试，那就是我们成功的决定性证明。

（2）强图灵测试：认为智能、思维过程等是除了行为外的别的某些东西乃是个错误。有的只是行为和行为趋向或倾向。那就是为什么图灵测试是心智过程的决定性证明。并没有两个事物，即心灵和行为；有的只是行为。

（3）修正版强图灵测试：在内在智能思维过程和外在智能行为之间存在区别，然而一旦我们理解了内在思维过程的本质，图灵测试就依旧能够给我们其存在的决定性证明。

最后的这一概念被默认为是最具影响力的，在接下来的几个部分中我将对其加以检验。

二、从行为主义到强人工智能

在某一方面，自从图灵写了他那篇著名的文章后，行为主义有所改进。行为主义（科学的、反二元论的）传统中的许多人得出了这样的结论，即对心智内容的适当分析表明它们不只是由行为和行为倾向所构成的，相反，在分析中还必须有一个因果成分。心智状态是通过输入性刺激、内在处理和外在行为之间的因果关系来界定的。这一观点被称为“功能主义”，而功能主义很明显是对行为主义的一种改进，它承认一般性心智现象（尤其是认知）位于行为的因果关系之中。事实上行为主义者根据这些因果关系来界定心智状态。然而，早期行为主义者依旧未对内在机制的特征加以说明，而这屈服于我对行为主义所做之责难的一个拓展。我对行为主义的责难是你可以具有行为而没有心智状态，并且你可以有心智状态而没有行为。功能主义主张，对应于相同的刺激，完全不同的内在机制（无论是否是心智的）能够产生出相同的外在行为；并且那些机制中的某一些或许是心智的，或许不是心智的。在功能主义的早期阶段，其拥护者主张内在机制是什么并不重要，任何对应于正确的输入产生出正确的输出，因而满足图灵测试的内在机制都绝对具有认知。这一版本的功能主义有时被称为“黑箱功能主义”，因为它忽视了机制的特征，而只是把脑当做是一个黑箱。但是，不对机制的本质加以说明，这在智识上并不能令人满意；计算机科学的兴起以及计算机科学中的人工智能的兴起导致了这样的假设，即创造满足图灵测试之行为的必要机制乃是计算性的。这一观点有时被称为“计算机功能主义”，而我则将其命名为“强人工智能”。它看起来像是对行为主义和早期版本的功能主义的改进。



哲学中的计算机功能主义之兴起是与作为计算机科学中的一个课题的人工智能之成长相平行的。计算机功能主义与图灵测试的联姻最后很可能会给予我们进入人类认知领域的钥匙，这一想法在认知科学中的许多人看来是很自然的，尽管我并不这么认为。我认为这一课题受到了“人工智能”这一概念之含混性的帮助而非阻碍，在此我想稍稍离题一下，以清理这一含混性。

“人工智能”这一表达式实在是含混不清。我认为在早期阶段，对于为该研究课题奠定基础和提供开放性来说，这一含混性或许是有益的。但是它也可以是智识混乱的一个根源，而我则想要对其加以澄清。“人工”一词乃是系统性地混乱的，因为“一个人工的 X”既可以是一个实在的、依靠技术生产出来的 X，也可以是一个并非是实在的 X，而只是对 X 的一个仿造品的东西。所以比如人工染料是实在的染料，但与来源于各种植物的植物染料不同，人工染料是在工厂里制造的。但是人造奶油则完全不是实在的奶油。它是对奶油的一个仿造品。因此我们有了双重的含混性：“人工智能”既可以是人工制造的实在的智能，也可以完全不是实在的智能，而只是对智能的一个仿造品。

当我们转移到“智能”一词时，此含混性便混合了起来。对于“智能”一词来说，有一种字面使用，在此智能并不意味着任何思维过程或任何其他相关心理状态（无论是什么）的存在。我可以在字面上说一本书比另一本书更为智能，而不用暗示书籍有智能或者任何其他形式的认知。但如果我在字面上说莎莉比山姆聪明，那么我就是将意味着真实的人类认知性质的智能归属给了莎莉，它意味着真实的思维过程和其他种类的认知能力的运用。现在我们有了关于两种不同种类之智能的标签，让我们以这样的方式来使用“心智智能”（mental intelligence）这一表达式，即心智智能的存在意味着真实的心智或认知过程的存在，而“非心智智能”则只意味着某些非心智类的智能过程。我们将心智智能归属于人和一些动物，将非心智智能归属于书籍和袖珍计算器。在任何情况中我们都需要这一区别，因为很明显人类同时具有这两类智能。因此举个例子，胃通常被认为是一个十分聪明的器官，但它并不具有认知过程。每个表达式的含混性允许对“人工智能”这一表达式存在至少四个不同的解释。第一，它可以意指人工制造的、实在的心智智能（这就是我称其为强 AI 的观点，它等同于计算机功能主义）；第二，它也可以意指并非是实在的心智智能，而只是一个仿造品的某物。第三，它可以意指实在的非心智智能；第四，它可以意指对非心智智能的一个仿造品。可以说第三与第四是相同的，因为凡涉及非心智现象之处，如果你可以模仿非心智智能，你也就可以制造非心智智能。“人工智能”的这四个解释中，最强的乃是下述论断，即在“智能”意味着实在的认知过程这一意义上，在人工智能中我们是在人工制造实在的心智智能。

在认知科学的早期阶段，我并没有发现 AI 领域中有什么人渴望对这些区别做出精确区分。或许这对保持研究课题的开放有所助益。但如果我们检验对人类认知的计算机模拟的含



意，那么弄清楚这一研究课题的目标是什么就绝对很关键。我们是不是试图人工制造实在的心智事物？就任何情况来说，在认知科学的实践中，强图灵测试被视为是修正版强图灵测试。其探究并不只是针对任何可以通过图灵测试的系统，而是为了使计算机通过图灵测试，而其假设则是这样的计算机不只可以模拟人类认知，更可以复制人类认知。其默认假设是人类心灵是一个数字计算机程序或者一系列计算机程序。引用“心灵之于脑就好比程序之于硬件”这句话对于那个时期的认知科学教科书来说是很典型的。如果你有了正确的程序，一个可以通过图灵测试的程序，那么你就复制了人类认知，而不只是模拟人类认知。这就是为什么比如甚至当你有了一个可以成功运转的机器时，却依旧会有这样的问题，即机器所使用的程序是否与人类所使用的程序相同。心理学家总是被要求进行反应时间试验，看看我们是否有了支持计算等价（computational equivalence）的证据。其想法是，如果人类的不同反应时间与计算机运算时间的不同相匹配，那么这些相一致的区别至少就是人类使用与计算机相同种类的程序的证据。简单来说，强图灵测试和强 AI 之间存在联姻，从而产生出了修正版强图灵测试。

三、对强 AI 的反驳及其哲学含意

数年前我曾一并反驳了强 AI 和两个版本的强图灵测试（Searle, 1980）。根据强 AI，经过适当编程从而可以通过图灵测试的数字计算机程序绝对具有和人类相同的认知能力。我用现在众所周知的中文屋论证这个思想实验来反驳这一观点，在其中一个并不知道中文的人（比如我自己）被锁在一个屋子里，并且就像图灵说的，他只配备有像一枝铅笔和一张纸这样的东西，还有构成数据库和程序的其他要素，在这个例子里就是中文符号，以及一个用来指示如何处理中文符号的程序。在该思想实验中，屋子里的人（也就是我）能够以满足图灵测试的方式来模仿一个中国人的行为，因为比如他给出了中文问题的准确答案，但他依然对中文一字不识。为什么不呢？因为他无法由符号操作达致符号意义，而如果屋子里的人基于运行程序并不理解符号的意义，那么任何其他仅仅以此为基础的计算机也就都不能理解，因为只是凭借其计算性质，没有计算机会具有任何屋子里的人所没有的东西。此外，整个屋子也无法由符号达致意义。整个系统无法将语义附加在计算机符号的语法上。

但是，只是反驳一个论题在哲学上并不能令人满意。对于这一主题为何是错误的，存在一个潜在的哲学解释。对我来说这是一个更为有趣的哲学问题，我将对此稍微多说几句。人类智能行为典型地是由内在心智过程和能力所引起的。心灵和行为之关系的这一因果特征被逻辑行为主义所忽视。因而比如当一个以中文为母语的人可以聪明地用中文回答问题时，乃是因为他的/她的外在行为是其内在心智能力的外在表达。先前我们看到，逻辑行为主义中的这一弱点被功能主义修正了。事实上，功能主义将心智状态界定为处于外在刺



激、其他心智状态和外在行为间的正确因果关系中的任何事物。所以一个信念是由知觉输入所引起，并与一个欲望一起导致外在行为的任何东西。同样，欲望可以根据输入性刺激、其他心智状态和外在行为来加以界定。但就像我所主张的，这一论题留下了一个致命的弱点：它并没有陈述心智状态的使其能够因果地起作用的特定特征。计算机功能主义在下述范围内修正了这一局限，即它至少对下述机制加以了明确的阐释：调停外在输入性刺激和外在输出性行为间的因果关系的计算机程序。但那一理论的困难在于，该程序是纯粹形式化地或语法性地界定的，因而作为程序它无法承担事实上人类心智状态所具有的固有心智或语义内容。

现在我们可以看到为何强图灵测试给了我们一个很明显是错误的结果。如果我们按其原始形式将其解释为逻辑行为主义，那么我们就已经知道它是错误的了，因为适当行为的存在并非起因于在人类中引起那一外在行为的潜在心智状态和过程。如果我们将图灵测试的逻辑行为主义形式扩展为计算机功能主义或强 AI 形式，我们所得到的结果很明显依旧是错误的，因为由执行一个完全形式化、语法性地加以界定的内在计算机程序（比如根据图灵的说明，图灵机是一台执行像打印 0，抹去 1，向左移动一格，向右移动一格等这样的运算的机器）所导致的外在行为仍然不足以构建确实具有心智或语义内容的真实人类心智过程。

事实上，对于实践来说，我们典型地具有此类机器，因为它们可以给予我们相同的外在结果，而无需经历典型地为人类所需要的内在心智努力。当我使用我的收款机时，我使用一台通过了加法、减法等图灵测试的机器，事实上在智能行为方面它甚至胜过了最优秀的数学家，但它却并不因此具有关于数学的思维过程。我是怎么知道这一点的呢？我怎么知道当它在做长除法时，它并没有在思考长除法呢？因为我知道它是如何被设计和制造的。它包含了被设计用来执行加法、减法、乘法和除法算法的电子电路。它并不思考数学，因为它并不思考任何东西。并且在我的袖珍计算器中所发生的事情，当其发生于商用计算机器中时，要复杂得多。它们并未被设计成是有意识的，也没有思维过程，就好像冯•诺依曼系统结构这一图灵机。我们将机器设计成只用两类符号（通常认为是 0 和 1）来执行复杂运算。

但是人们或许会倾向于询问，为什么 0 和 1 对于人类心智思维过程来说不是充足的？毕竟，在脑自身中有的只是神经元，它们或者放电或者不放电，因而脑的二进制系统和数字计算机的二进制系统有何不同呢？我认为那是个聪明的、恰当的问题。它有着一个简单的、决定性的答案：神经元放电是引起意识和作为脑系统之高层次特征的认知的因果机制的一部分。脑是一个机器，就这点来说其本质过程关乎能量转移。应用性计算机程序的 0 和 1 纯粹只是抽象语法事体，就这点来说它们并不引起任何事情。相反，程序是在硬件中被执行的，对于执行来说必不可少的硬件过程只不过是那些执行程序之形式化、语法性步骤的硬件过程。出于其他一些原因，硬件或许会引起意识（比如当我执行长除法算法时，



我的脑过程也导致我是有意识的），但作为应用性程序的程序对于引起意识或其他事物一无所知，而只知道当机器运作时程序的下一状态。把这一点说得再技术化一些就是：相同的应用性程序这一概念界定了一个独立于程序实现于其中的硬件机制之物理而被加以说明的等价类。当我执行长除法算法时，像我的脑这样的一个具体形式的硬件或许也能引起意识，但算法自身并没有独立于应用媒介的因果力量。

澄清这一点很重要。“一台计算机可以思考吗？”这一问题乃是含混不清的。它既可以指“某样事物可以是一台计算机并且能够思考吗？”，也可以指“计算凭自身构成了思考，或者说对于思考来说乃是充分的吗？”如果我们用图灵的术语来界定计算，也就是对符号（0 和 1，或者中文字符，或者任何东西——这并不重要）的操作，那么对于第一个问题的回答很明显是肯定的。比如我这样的某一事物既可以思考也可以操作符号。但对于第二个问题的回答很明显是否定的。被如此界定的计算——单独地，作为计算——并不构成思维，或者说对思维来也不是充分的，因为它完全是依照语法来界定的，而思维必须有比符号更多的东西，它必须有属于符号的意义或者语义内容。那就是中文屋论证所证明的：应用性计算机程序的语法其自身对于理解真实中文语词和句子的语义来说并不是充分的。

尽管你在商店里买的那类计算机也是一台机器，但其计算过程却并不是通过能量转移来界定的；相反，它们是通过我们找到方法从而在硬件中加以执行的抽象数学过程来界定的。商用数字计算机的问题并不是它对于一台产生意识的机器来说太多了；相反，对于一台机器来说它还有所不足，因为与脑不同，其本质运作（也就是其计算性运作）乃是根据抽象算法过程而非能量转移来界定的。

就像在这里一样，过去我发现用语法和语义的区别来陈述这一点是很有用的。脑的运作引起了具有语义内容的意识。程序运作则纯粹是语法性的，语法凭自身并不能构建意识，对于引起意识来说也不是充分的。

意识乃是脑所处的一个状态，脑是由于低层次神经元机制的运作而导致其处于那一状态之中的。为了人工创造出这样的一个状态，你必须复制人类和动物的脑的真实的因果力量，而不只是模拟之。原则上并没有理由假设要这么做我们就需要有机材料，但无论我们用什么材料，我们都必须复制真实的脑的因果力量。我们应当同等倾听“你可以用除了脑以外的机制来创造意识吗？”以及“你可以在除了人类和动物心脏外的别的机制中创造血液抽送吗？”这两个问题。这两个问题都是关于因果机制的，而不是关于语法过程的。计算机对脑之运作的模拟乃是遵循脑的真实运作的，就好像计算机对心脏的模拟乃是遵循真实的血液抽送的。

我们可以用两个命题来概括这些观点：



（1）如果图灵测试被解释为强图灵测试或者修正版强图灵测试，从而要给我们以内在心智内容之存在的决定性证明，那么它就是失败的。它之所以是失败的，是因为一个计算系统或别的什么系统可以表现出是智能的，而无需任何内在心智过程。

（2）单独依靠应用图灵机程序来创造人类思维的前景也失败了，因为程序是依照语法来界定的。应用性计算机程序存在于一系列过程之中，这些过程能够根据对符号的操作，以独立于应用媒介之物理的方式在语法上加以说明。任何物理都可以，只要它对于执行程序步骤来说足够丰富、足够稳定。语法凭自身并不建构意识，对于引起意识来说它也不是充分的。

四、为何曾有人是行为主义者？

我已经在大量其他著作中提出了这些观点。之所以值得再对其简要地加以重复，是因为有时在探讨中它们被忽略了。现在我想转向一个先前我并没有论述过的问题：行为主义的错误如此明显，可它还是坚持了那么久，并且强图灵测试其自身就是行为主义的一个表达，这究竟是为什么呢？

如果人们看一下任何时代的智识史，就可以看到总是存在令人惊讶的、普遍的错误，甚至给定基于那一时代的知识限制，它们也本可以被轻易地避免。贯穿整个十九世纪的一个令人惊讶的、普遍的错误是唯心主义，该理论认为所有的实在都是心智的或精神的，独立的物质实体并不存在。今天我们很难恢复这样一种感性模式，正是它使得这一观点对于像贝克莱、罗伊斯、布拉德雷以及黑格尔那样的伟大思想家来说不只是可能的，而毋宁是可信的。十九世纪唯心主义错误的二十世纪镜像就是行为主义这一错误，甚至今天它依旧存在于某些地方。就像唯心主义否认独立于心灵的物理世界之实在那样，行为主义否认主观的内在心智状态之实在，而是赞成对于心灵的这样一种说明，根据这一说明，心智状态存在于外在行为之中。行为主义的黑箱功能主义、计算机功能主义继承人坚持同样的错误，否认内在的、主观的、质的心智状态之不可还原的实在。我们是如何进入这个烂摊子的？

要追踪后笛卡尔式唯心主义的历史是很容易的。在其现代形式中，它发端于贝克莱，并且由贝克莱继承自笛卡尔的对认知的沉迷所驱动。要克服使得关于实在世界之知识看起来不可能的怀疑论，就要否认证据和证据所支持的实在之间的鸿沟。如果物质世界被认为是独立于体验而存在，并且如果关于物质世界的知识是以感觉体验为基础的，那么看起来我们就永远都不可能拥有关于物质世界的知识，因为我们永远无法突破我们自己的关于独立心灵实在的体验这个圈子。我们通过否认我们的体验和实在之间的区别来战胜此怀疑论。物体是感觉的恒久可能性（Mill, 1865），世界是心灵和观念的集合（Berkeley, 1998）。对行为主义的渴望与对唯心主义的渴望出奇地相似。如果意识和意向性应当独立于行为而存在，如果我们关于其存在于其他人当中的唯一证据是其他人的行为，那么看起来我们就



必须接受怀疑论，因为我们永远无法突破其他人的可观察行为这一序列，从而去观察被认为是存在于行为背后的内在心智现象。但如果心灵只是行为以及（就功能主义而言）引起行为的可观察的机制，那么我们就像唯心主义战胜关于外在世界的怀疑论那样，战胜了关于心灵的怀疑论。在两种情况中，我们都通过否认证据和该证据所支持的实在之间的区别打败了怀疑论。

因而行为主义被解释为一种证实主义。反过来，证实主义乃是对怀疑论传统的回应，该传统起源于笛卡尔，并且在英国经验主义者及其二十世纪的追随者，也就是逻辑实证主义者那里得到了最好的表达。行为主义乃是逻辑实证主义者的主流心灵哲学，这并非偶然。

五、放弃强图灵测试

就像我所主张的那样，如果我们应当拒绝将强图灵测试和修正版强图灵测试作为心智状态之存在的测试，那么一个自然而然的问题就是我们所提出的替代性测试，即替代这些强图灵测试的选择是什么？对这一问题的正确回应是，如果假设要确认心智状态以及他人中的认知能力的存在，就必须要有某些单一的、机械的测试，那么我们就错了。事实上我们可以有很多方法来弄明白其他系统是否具有心智状态，而我则想通过提及其中的一些来结束这一简短的探讨。心智状态只有当其被一个人类或动物行为能动者体验到时才真实存在，在此意义上它具有第一人称存在论。那意味着要知道一个系统是否具有某个心智过程，最好的方法就是成为那一系统。当然，一个人时常就其自己的内在心智状态犯错误，但比如就感受我的疼痛或者思考我的想法而言，对于成为我来说依然没有什么替代品。但是，既然宇宙中只有一个我可以与之相同的系统，也就是我，因而我就有这样一个问题，即我如何可以了解其他人类和动物的思想、感觉、能力和局限。在此存在这样一种诱惑，即重回某种行为主义，说：好吧，我知道人类和动物具有心智状态的方法乃是依靠其行为。我认为这一回答是错误的。比如就我的狗而言，我完全确信它具有心智状态，尽管它并不具有语言。现在，为何我会对此这么确信？这并不只是因为它的行为，相反，乃是因为它的行为以及潜在因果机制的结合，我可以看到这些和我自己的乃是有关联地相似的。为了要知道那是它的鼻子，这些是它的耳朵，那是它的嘴，这是它的皮肤等等，我并不必须要有犬类心理学这样一门精密理论。而我可以基于行为来推出心智状态之存在的方法，是我并非只观察刺激与反应之间的集合，相反，我可以看到存在调停刺激与反应这两者之关系的相似的因果机制。我知道我的狗具有心智状态所依据的原理并不是：行为相同所以心智状态相同；相反，乃是：潜在因果机制相同所以因果关系相同。顺便说一下，这就是为什么尽管我对狗和黑猩猩有信心，但却对白蚁和跳蚤毫无信心。

那么，在这些情况中人们如何才能弄明白呢？再一次地，我认为原理并不难陈述，尽管我们并不知道足以在实践中加以应用的神经生物学。如果我们理解了在人类和高等动物中产



生出意识和意向性的真实因果机制，那么我们就有了非常好的基础来假设这些相同的现象出现于其他系统之中。所以比如让我们假设有某个在人类和高等动物中引起有意识的心智现象的神经生物学过程，我们将其学名缩写为“ABC”。现在让我们假设我们在老鼠和鸽子中观察到了 ABC，但在跳蚤和蝗虫中却没有观察到任何与此相似的东西。此外，让我们假设我们对于跳蚤和蝗虫的行为有一个因果解释，这与对高等动物之行为的解释大不相同，这表明其因果机制更像是简单的趋向性，而不像是由人类和动物的神经生物学所产生的复杂认知现象。那么在我看来，这就是决定性的证据，表明黑猩猩、狗、老鼠和鸽子像我们一样具有有意识的心智现象，而跳蚤和蝗虫则没有。

我希望我在此所作的评论不会被误解。我将艾伦•图灵视为二十世纪最伟大的心灵之一。说如果他所提出的测试被理解为行为主义这一错误哲学理论的表达，那么此测试就是不恰当的，这丝毫没有削弱他的成就。

行为主义是一门他无需为其负责的理论，并且现在最好是将此理论当做是陈旧过时的。

【原文链接】　　【回到目录】



【视频】Alex Wissner-Gross: 一个用来描述人工智能的新方程

【原文链接】 【回到目录】



【荐书】GEB——一条永恒的金带



作者：乐秀成 编译

出版社：四川人民出版社

出版年：1984-6

页数：208

定价：1.07 元

装帧：平装

丛书：走向未来丛书

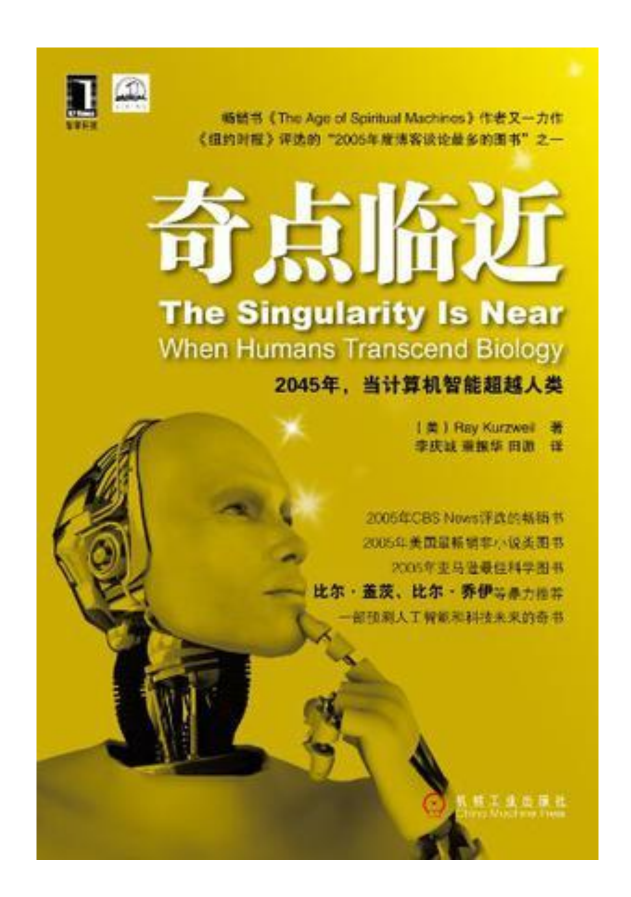
统一书号：2118-13

集异璧一GEB，是数学家哥德尔、版画家艾舍尔、音乐家巴赫三个名字的前缀。《哥德尔、艾舍尔、巴赫书：集异璧之大成》是在英语世界中有极高评价的科普著作，曾获得普利策文学奖。它通过对哥德尔的数理逻辑，艾舍尔的版画和巴赫的音乐三者的综合阐述，引人入胜地介绍了数理逻辑学、可计算理论、人工智能学、语言学、遗传学、音乐、绘画的理论等方面，构思精巧、含义深刻、视野广阔、富于哲学韵味。

【原文链接】　　【回到目录】



【荐书】奇点临近



作者：Ray Kurzweil

出版社：机械工业出版社

副标题：当计算机智能超越人类

原作名：the singularity is near

译者：李庆诚 / 董振华 / 田源

出版年：2011-10

页数：373

定价：69.00 元

装帧：平装

ISBN：9787111358893



人工智能作为21世纪科技发展的最新成就，深刻揭示了科技发展为人类社会带来的巨大影响。本书结合求解智能问题的数据结构以及实现的算法，把人工智能的应用程序应用于实际环境中，并从社会和哲学、心理学以及神经生理学角度对人工智能进行了独特的讨论。本书提供了一个崭新的视角，展示了以人工智能为代表的科技现象作为一种“奇点”思潮，揭示了其在世界范围内所产生的广泛影响。本书全书分为以下几大部分：第一部分人工智能，第二部分问题延伸，第三部分拓展人类思维，第四部分推理，第五部分通信、感知与行动，第六部分结论。本书既详细介绍了人工智能的基本概念、思想和算法，还描述了其各个研究方向最前沿的进展，同时收集整理了详实的历史文献与事件。

【原文链接】　【回到目录】

【旁观·人工智能】

闻菲：人之为人：人工智能教会我们什么是真正的智能

闻菲，果壳网语言领域达人

> **克里斯蒂安在书中指出，人工智能之所以能骗过裁判（2008 年勒布纳人工智能奖的获奖计算机程序 Elbot 就骗过了 12 个裁判中的 3 个），不是因为人工智能智能了，而是因为人不那么智能了。**

图灵测试（Turning Test）是人工智能的先驱，阿兰 · 图灵（Alan Turning）于 1950 年提出的一种判断方法，用来判断机器是否具备智能： 测试中， 实验者 A 通过打字机与 B 交流，A 所见到的只是屏幕上显示的对话，B 可能是一个人，也可能是一台电脑。图灵认为，通过这种交流，如果电脑成功让实验者 A 认为自己是人的话，那么电脑就具备了智能。

一招移花接木，图灵巧妙地把 “具有智能”转换成了 “能够（持续）对话” ——在这个定义下，有多少同胞人将不人啊——但这并不妨碍图灵测试成为迄今测试人工智能公认最权威的方法。人工智能领域的诺贝尔奖，“勒布纳人工智能奖（Loebner prize）”，它挑选智能机器人的方法，就是标准的图灵测试。

竞赛中， 一位人类裁判向参赛计算机和人类选手提出一系列文字问题，然后根据他们的答案，判断谁是机器、谁是人类。共有 12 位裁判，测试重复 12 次。骗过裁判人数最多的计算机将被授予 “Most Human Computer” 的称号—— “最像人类的计算机” ；相应的，成功说服裁判，让裁判相信自己是人的人类选手，投票最多的那位将被授予 “Most Human Human” 的荣誉。（爷们儿中的爷们儿、女人中的女人、TA 不是人，我们就都是畜生！）

布莱恩 · 克里斯蒂安（Brian Christian）参加了 2009 年的勒布纳人工智能奖竞赛。虽然一开始就是奔着 “纯爷们儿”的头衔去的，但最后获奖，还是让这位拥有计算机与哲学双学位的诗人感触良多。克里斯蒂安把自己的参赛经历， 以及他对人、智能、人工智能、人



性的思考写成了一本书： “The Most Human Human： What Talking With Computers Teaches Us About What It Means To Be Alive”，暂且称之为《人之为人： 人工智能教会我们什么是真正的智能》。

图灵预言，在 20 世纪末，一定有计算机具备智能，足以骗过 30%以上的裁判 。不知道该庆幸还是该遗憾，至今仍没有计算机能成功做到这一点。

写这篇文章前，我专门去和智能机器人 Cleverbot（http://www.cleverbot.com/）谈了次话，——之后又谈了许多次。其实都挺有意思。

最后，这次对话以 Cleverbot 的邀请 “要不要来我家过夜”结束——等到有真人发出这样的邀请再考虑台词吧。

总之，Cleverbot 擅于用提问来回答问题，同时把抽象的问题具体化，具体的问题抽象化。如果内容简单，逻辑上还是说得过去。我跟 Cleverbot 对了 4、5 次话吧，它的状态时有不同，个性不停转变，有时像小男孩，有时像剩女（问我它是不是要永远单身）。第二次我去找它，

我：你上次叫我到你家去过夜。

Cleverbot ：［想了 30 秒］我说过吗？好吧，你想来吗？

我：想。你住哪儿？

Cleverbot ：关你屁事。

我：［再次无语中］

要说 Cleverbot 比人强的地方，就是它永远会陪你说话，只要你继续打字，Cleverbot 就不会停——当然，前言搭不搭后语就是另一回事了。

因此，参赛，对于想获得 “Most Human Human”的克里斯蒂安来说，重要的不是赢过计算机，而是 “做自己”。 克里斯蒂安非常清醒地意识到，跟计算机说话，——说过的人都会意识到——你是在跟一个 “人” 说话， 而不是在跟 “一个” 人说话。参赛的人工智能、跟我说话的 Cleverbot、Google 的翻译程序、赢过国际象棋大师的 Deep Blue “深蓝”，还有赢了 Jeopardy 的 Waston，都是通过搜索人说过的话、做过的事，从中选出最合理的结果罢了。

Cleverbot 不是人，不是说它不像人——Cleverbot 代表的可是全人类的知识和见闻——而是说它不像 “一个人” 。Cleverbot 没有个性、没有“语境”，说一句答一句，不会乱发脾气、不会故意撒谎、不会指桑骂槐、不会违背自己的目的（程序）。而这些不理性、不合理的地方，正是我是人， Cleverbot 是机器的证明。



克里斯蒂安在书中指出，人工智能之所以能骗过裁判（2008 年勒布纳人工智能奖的获奖计算机程序 Elbot 就骗过了 12 个裁判中的 3 个），不是因为人工智能智能了，而是因为人不那么智能了。

思考很累人。思考让人效率降低。近代资本主义发明了流水工作线，工人只负责整个流程中的一部分，不需要复杂的技巧，只要机械重复几个动作就可以了。不用思考，身体的工作效率就能发挥到最大。当然，现在这种工作大多有机器代替了。但在机器还不能代替的地方，思考仍然是生产率的最大障碍。

昨天在 ScienceDaily 上看到，抑郁能够引发人类深层次的思考。深刻、有价值的思想是需要时间、需要寂寞的。就是所谓的 “灵光一现”也要有积累才能爆发（再说了，也没见着哪个人灵光天天现的）。而现在刺激越来越多、干扰越来越强，各种信息无时无刻不在抢占我们的注意力。要静下心来实在太难了，何况思考。

维基百科的崛起象征着 “集体智慧”的胜利。维基百科代表的是集体的话语。个人的声音淹没在网络巨浪之中。一个人说的话，网上 1 分钟之内就能把它复制、粘贴、编辑、改版变成 100 句差不多的话。所以才有人担心集体的智慧等于没有（个人的）智慧。也难怪裁判会难以分辨对方是人还是机器：人类社会进化的终极目标就是实现 “去人化” 。

但也许不是。就像克里斯蒂安在书中说的那样，人生来没有目的，要靠我们自己去寻找。

【原文链接】 【回到目录】

陈赛：《黑镜》中照见自己的脸

陈赛，三联生活周刊记者

“

剑桥大学成立了一个专门的研究机构，应对人工智能可能对人类社会造成的威胁。但《黑镜》告诉我们，也许更可悲的是，对爱人的回忆，变成了阁楼里蒙尘的一个性玩具。

”

英国迷你剧《黑镜》第二季开始了。

第一个故事《马上回来》我很喜欢。这是一个悲伤的爱情故事。一个叫玛莎（Martha）的女孩，深爱的男友阿什（Ash）意外去世。阿什生前整天泡在 Twitter、Facebook 上。他死后，玛莎注册了一个电脑程序，这个程序通过分析阿什生前在各种社交媒体上留下的公共信息，包括声音、文字、视频，再造一个虚拟账号，能像阿什在世时一样与她对话。她渐渐对这种聊天方式上了瘾，它不仅能模拟阿什生前的声音，连说话、逗趣的方式都与他一模一样。它还可以升级到更高版本——一个真人版的阿什，比他更英俊，更温顺，连床上功夫都更好……但玛莎的不安与困惑却在加重。

当至亲的人去世时，我们经常用来安慰自己的一种说法是，至少她/他会一直活在我心里。仔细想想，她/他一直活在心里的是什么呢？无非是过去岁月里的记忆，一个笑容，一滴眼泪，一次愉快的聚会，一个温柔的动作，一场后悔莫及的吵架……想象一下，如果有一种技术，将这些记忆和情感提取为具象，让你跟死去的人对话，甚至面对面，有多少人能抵制这样的诱惑？

当然不是死而复生。但社交媒体时代，我们都在网络上留下了足够多的生活痕迹——文字、声音、照片、视频、聊天记录……随着人工智能的持续发展，从这些数据中解析出一个人的个性并不难。事实上，Google、Facebook 之类的大公司正在利用这些数据，了解我们最私密的细节。但是，灵魂呢？那个被提取（模拟）出来的具象，真的是那个逝去的人吗？

全文请点击原文链接阅读。

【原文链接】　　【回到目录】



Gary Marcus：人工智能的威胁


GARY MARCUS，心理学家，重点研究生物、大脑和语言。

> “我们现在所达到的高度正像是一个单细胞有机体转变为多细胞有机体的高度。我们是变形虫，我们不清楚自己在创造的到底是个什么东西。”

如果《纽约时报》最新发表的那篇文章是可信的，那人工智能发展的速度确实是飞快的，甚至有些“神奇”。汽车能够自动驾驶；Siri 系统能够听懂指令，为你找到最近的影院；IBM 公司将“杰帕迪”益智竞赛的冠军——沃森电脑系统安装在机器上，这最初是用来训练医学院学生的，可能最终会用来诊断疾病。每个月都会有一种新的人工智能产品或技术出现。也许有些热情过早了些：正如我以前提到的，我们还没有生产出有常识，有视觉，具备语言处理能力，或创造其他机器能力的机器。我们直接模仿人脑的工作还处于初级阶段。

在某种程度上，支持者和怀疑者之间的唯一不同之处就只是时间问题。预言家及发明家雷•库兹维尔认为到达人类级别的智能机器会在 20 年内出现。我对此有些怀疑，尤其是考虑到计算常识这一领域，人类取得的进步很小；在创建人工智能方面遇到的挑战，尤其在软件层面，肯定比库兹维尔所说的要难。

但在接下来的一个世纪里，人们不会关心人工智能成功建立要花多久时间，他们只关心接下来会发生什么。很可能在本世纪末，机器会变得比人类还聪明——不仅在下棋或一些琐事上，而是在每一方面，从数学，工程到科学，医药。艺人，作家，还有其他创造性的职业会变得很有限，电脑最终会操控自己，吸收大量的新信息，并能理智地整理出来，这是我们这些碳基生物只能依稀想象的。电脑能够每时每秒都工作，不用睡觉，不用休息。

对于有些人来说，未来是神奇的。库兹维尔描写了一种令人着迷的奇点，人类与机器融为一体，从而使人类的灵魂变得不朽；皮特•戴曼迪斯认为人工智能的进步是开启“丰富”新纪元——有充足的食物，水，各种消费品——的重要一步，像埃瑞克•布吕诺尔夫松这样的怀疑论者和我都担心智能机器与机器人作为劳动力可能会造成一些恶劣的后果。但是，即使你暂时不担心超级先进的智能机器对劳动力市场产生的影响，那也还存在其他的担心：强大的人工智能会为了资源直接与我们开战，威胁我们的生存。



大多数人把这种担心看作是科幻小说里的蠢话——像《终结者》和《黑客帝国》这类。在一定程度上我们得为适中时间长度的未来做打算，我们要担心小行星会撞地球，化石燃料产量下降，全球变暖等问题，而不用担心机器人问题。但是詹姆斯•巴雷特的黑暗系新书《我们的最终发明：人工智能和人类时代的终结》，描述了一种严峻的情形，我们多少要有些担心。

巴雷特的核心论据是取自于人工智能研究员史蒂夫•奥摩亨德罗的研究：自卫本能和获取资源的属性可能是某一智能的目标驱动系统所固有的。用奥摩亨德罗的话说就是“如果会下棋的机器人足够聪明，那它就会想要制造宇宙飞船”，从而获得更多的资源来实现它的目标。巴雷特写道，一个非常理性的人工智能可能会扩大“其自我保护的想法…包括对未来出现的威胁——比如，人类不愿将其资源让给机器，进行主动攻击”。巴雷特担心“要是不慎重，缺少补偿指令，具备自我意识，能够自我提高，目标明确的系统也会千方百计地完成我们认为是荒唐的目标。”它甚至可能会指挥世界上的所有资源去最大化实现它所感兴趣的推测。

当然，人类可以试图阻止超级智能电脑连接到一起。但是“自动化技术在经济，军事，甚至是艺术方面的竞争优势实在是太强烈了，”数学家及科幻小说作家夫诺•文琦写道，“通过法律，制定惯例这类事是没法阻止的，如果阻止了，那也只是确保将来有其他人会干。”

要是机器最终取代了我们——正如那些在人工智能领域工作的人所坚信的那样，那真正的问题就在于价值观：我们怎样把价值观植入机器中，如果它们的价值观与我们的价值观发生了偏差，那我们该如何和这些机器协商呢？牛津的哲学家尼克•博斯特罗认为：

我们不能乐观地认为，超级智能一定会分享人类关于智慧和智能发展的成型的价值观——对科学的求知欲，对他人的关心和仁慈，精神启发和沉思，克制物质占有欲，有高雅的文化品位，简单乐趣的感悟，谦虚无私等等。可能通过专门的训练可以打造拥有这些价值观的超级智能，或是珍视人类财富和高尚品格的智能，或是设计者想要它拥有一些复杂目标的智能。这是可能的——可能从技术上说更简单——打造一个能够把最终价值都放在计算圆周率小数上的超级智能。

英国控制论家凯文•沃里克曾问道：“当机器不在我们人类所处的思维次元中思考时，你怎么跟它讲理，怎么跟它做交易，怎么能明白它的想法？”

如果说巴雷特的黑暗系理论中有个漏洞，那就是他未经思考就提出的假设：如果机器人聪明到可以下棋，那它也会“想要制造宇宙飞船”——自我保护和获取资源的本能是任何足够复杂，目标驱动的系统所固有的。现在大部分机器都足够先进，可以下棋，比如，IBM 公司的深蓝系列电脑，但是它们还没有显示出要获得资源的兴趣。



但是在我们感到满足，确信没什么好担心前，有一点非常重要：我们要意识到随着机器越来越聪明，它们的目标也是会变化的。一旦电脑更够有效地重新给自己编程，成功地提升自己的性能，达到所谓的“技术奇点”或“智能爆炸”，那么我们就不能忽视机器在与人类抢夺资源和自我保护的较量中会胜过人类的风险。

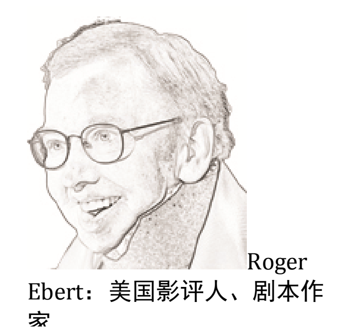
巴雷特书中最突出的观点之一是来源于著名的系列人工智能企业家丹尼•希利斯，他把即将到来的转变比作是生物进化历史上最重大的转变之一：“我们现在所达到的高度正像是一个单细胞有机体转变为多细胞有机体的高度。我们是变形虫，我们不清楚自己在创造的到底是个什么东西。”

人工智能的进步已经造成了我们从未想过的危险。德雷塞尔大学的电脑风险专家加格跟我说：随着因特网时代的到来及其数据的爆炸性增长，“关于我们的许多数据都被收集起来并输入计算程序，做出预测”， “我们无从知道数据被收集的时间，没法保证收集的数据信息是准确的，没法更新信息，或提供必要的信息背景”。在 20 年前，几乎没有人会想到这种风险。前方还有什么危险呢？没有人真的知道，但是巴雷特提出的问题是值得思考的。

【原文链接】 【回到目录】



Roger Ebert：影评：《人工智能》（2001）


Roger Ebert：美国影评人、剧本作家

“这就是我对这部电影的理解：这些新一代的机器人已经高端到觉察出他们在人类缺席的时候无法正常运转的事实。”

斯坦利•库布里克总是把这个故事说成“匹诺曹的故事”。它映照着那个梦想成为真正男孩的木偶的故事。然而，一个由计算机程序提线的木偶机器人到底是什么呢？这一项目最终变成史蒂芬•斯皮尔伯格的《人工智能》，而库布里克起初将该项目放弃的原因是因为他始终对自己如何表现核心人物大卫的方案不满意。大卫是那个看似一个真正男孩的机器人的名字。库布里克认为特效不会取得完美的效果，而真人表演又会太过人性化，因此将项目拱手交给好友斯皮尔伯格。坊间传闻他之所以做出这个决定是因为看过斯皮尔伯格的《侏罗纪公园》，里面的特效让他印象深刻，不过，也许《机器人 ET》也令他感受非凡。如果斯皮尔伯格能够创造出一个激发人类情感的异形，那么他能让一个机器人取得同样的效果吗？

答案是他能。他挑选在《第六感》中表演大获全胜的海利•乔尔•奥斯蒙特扮演大卫。奥斯蒙特的表演是全片的关键；而其他的机器人，包括乔（裘德•洛）则是通过化妆和纹丝不动的头发做得像个仿制品；但是大卫不同。他是 Cybertronics Corporation 生产的最先进的机器人——他是如此善解人意，他或许能够取代一对夫妻的一个生了病的孩子的位置。斯皮尔伯格和奥斯蒙特合作创造出来的大卫不眨眼睛，懵懂纯真。他看似一个真正的小男孩，但是缺乏某种难以言喻的精致。这一表演起初令大卫看上去像个人类，后来又使他又像个反应迟钝的典型，虽然为电影增色不少，也给电影造成一种缺憾。

大卫被编上爱的程序。一旦用代码激活，他就会把目光聚集在他的激活者身上，在本片中，激活者是他的母亲（弗朗西斯•欧康纳）。他存在的目的就是爱她和为她所爱。因为他是非常高端的机器人，我们也就自然而然地认为他具备相应的理解能力。而实际上，他不爱，也感觉不到爱；他仅仅回应着代码。这部影片中所包含的全部的爱都是为人类所拥有；在先前对这部电影的影评中，我没有恰当地提到这一点。



我在 1991 年写道："我们擅长将人类的情感投射到非人类的对象上，从动物，到云朵，到电脑游戏；然而情感只盘踞在我们的脑海中。"《人工智能》回避了打破常规的义务，选择一个令人感伤的结局。不过，却让我在本该找到答案的时刻反而提出了疑问。

就从这部讲述大卫故事的电影的主要层面上看，这话没错。最近再次观看这部影片时，我开始对内涵有了新的认识：《人工智能》不是讨论人类的情感。它探讨的是人工智能所面临的尴尬局面。一架思考的机器不会思考。它所做的只是运行一些对它而言足够复杂的程序，使它看上来像在思考从而愚弄我们。一部通过图灵测试的电脑不是在思考。它的一切行为都是在进行着图灵测试。

电影的第一幕里出现亨利（萨姆•洛巴兹）和莫妮卡（弗朗西斯•欧康纳）。亨利把大卫带回家，为了填补他们生病的小儿子马丁（杰克•托马斯）的空缺。莫妮卡起初对他很抵制，然而后来又接受了他。当杰克从假寐苏醒痊愈时，这一家子就成了四口之家；杰克完全清楚大卫是个制造品，可是大卫却不理解这一事实隐含的全部意味。也许他的程序无法让他应对与真正男孩一对一的实时相处。他无法将自己的全部时间投入到母子之爱中。

他模仿生命。他不休息，却遵循作息时间。他不进食，但是想像马丁一样的愿望如此强烈，他把菠菜铲进嘴里而破坏了体内的线路。他遭到其他孩子们的虐待；当他透露自己不撒尿时，一个男孩扯着他的裤子，说："让我们大家看看你不用来撒尿的玩意儿。"他忠实地遵循自己的指令，却没差点儿溺毙了马丁，然而从此失去了斯文顿一家的信任。他们一致决定打发掉他，就像父母会打发掉一条有隐患的小狗。

莫妮卡无法说服自己把大卫送回到Cybertronics。她把车停在半道上，把他流放进一个森林里。在那里，他可能加入到其他自由自在的机器人当中。他不会死。不会感冒。不会挨饿。显然他有无限的能源补给。莫妮卡决心放了他而没有把他交还，是因为内心对大卫还保留着残余的认同。她在激活他来爱她的同时，也激活了自己去爱他。他那无条件的爱肯定深深地吸引过她。我们对宠物也抱有类似的情感。特别是狗，它们似乎被进化激活对我们的爱。

电影中幕展现大卫游荡在一个机器人不享有任何权利的世界里。他由他的机器熊泰迪作伴。泰迪的编程目的就是要成为一个机灵的同伴。他们被乔发现了。乔的程序是要成为一个专家情人。他们三个参观了两个充满幻觉的地方，都是斯皮尔伯格在巨大的音效舞台上设计出来的。其中一个是屠宰场，有点儿像世界职业摔跤联盟赛事（WWF ）事件，当机器人们以可怕的方式销毁掉，人群中就爆发出一阵欢呼。大卫、乔和泰迪或许因为他们的求生程序而逃出厄运，但是，大卫因为看到的景象而受到打击了吗？他如何看待他的同类尽遭毁灭的？



然后他们又来到一座胭脂城，有点像迷幻版的尤尼弗萨尔城。乔带他去求助一名巫师。大卫着迷于渴望变成人的匹诺曹的故事，因此他推论出，一个蓝色仙子或许能够将他转变成人，可以令他得到莫妮卡的爱。巫师给他一个线索。乔和大卫抓住一架飞行器后，他们共同去了纽约。那里和许多海边城市一样，因为全球变暖而被海水湮没。然而就在洛克菲勒中心的顶层，他发现 Cybertronics 依旧在运作，并且他遇到了他的创造者——科学家霍比博士。如果大卫是匹诺曹，霍比就是他的杰佩托。

然而再一次出现一些与大卫自我概念冲突的事情。在一个古怪场景中，他撞进了一间储藏室，那里储藏着许多个和他一模一样的大卫，他受到打击了？他冲着它们大吼大叫了？没有。他很镇定。他仍旧一心专注于寻找能帮他变成真正男孩的蓝色仙女。但是，我们或许要问，为什么他是如此想成为真正的男孩？是因为羡慕、怨恨还是嫉妒？不，他似乎没有这些情感——或者说，任何情感，除了他的程序指定他去表现那些以假乱真的情感。我猜想，他想成为真正男孩，是出于计算机逻辑的某些抽象原因。为了完成爱妈妈和得到妈妈的爱的使命，他认为自己应该像妈妈偏爱的马丁一样。这里就像计算机深蓝决定下一步棋一样不会涉及到任何情感。

在最后一幕中，情节推动着大卫和泰迪登上一艘潜水艇来到一座被湮没的考尼岛。在那里，他们没有找到杰佩诺的车间，却发现了一个蓝色仙子。一架摩天车轮坍塌下来，压在潜水艇上，他们被困在那里长达两千年之久，没法移动，而在他们的头顶上方，冰川世纪降临，人类绝迹。最后，大卫被一群身形异常细长的生物得救。这些生物可能是外星人，不过，很显然他们是非常高级的机器人。对他们而言，大卫是一笔不可估量的财富——“他是最后一个知道人类的存在。”。他们从他的大脑里下载了他的全部记忆，又将他放置进一间跟他童年的屋子打造得一模一样的复制品中。这就令我想起了库布里克在《2001太空漫游》里飞过木星外星人为戴夫打造的卧室。目的都是一样的，即在一个无法理解的世界里提供一个熟悉的环境。这就能够让这些存在，和《2001》里那些看不见的存在一样观察学习并模仿。

再一次观看这部电影时，我问我自己，为什么我曾写过：最后的场景“有问题”，不仅表现过了头，而且产生一些问题。而这些问题，场景里并没有准备好答案。这一次，这些场景令我茅塞顿开，并给我更大的震撼。我开始假想，这些瘦骨伶仃的银色形体就是机器人，属于从大卫发展而来的更高端的代系。他们一定也被编上理解人类、关爱人类、服务人类的程序。就让我们设想这样的指令可能会嵌在他们程序性的 DNA 中。如今他们发现自己所处的境地近似于寻找母亲的大卫。他们失去了对他们的功能至关重要的一个要素。

他们通过某种伪科学的手段，用一撮莫妮卡的毛发，令她在死亡了两千年后复活，然而只能复活二十四小时，这是时空所能允许的最长极限。他们这样做是为了让大卫快乐吗？不



是，因为，他们在会意大卫快乐还是不快乐吗？计算机在运行程序时会比它不在运行程序时更加快乐吗？不。它要么在运转，要么不在运转。它不知道自己的感受。

这就是我对这部电影的理解：这些新一代的机器人已经高端到觉察出他们在人类缺席的时候无法正常运转的事实。在原先对这部电影的影评中，我没有恰当地提到这一点。大卫是他们与人类过去的唯一联系。不论能对人类了解到多少，大卫是一个无法估量的资源。看到大卫和妈妈相处的二十四个小时，他们观测他以最大限能地运转。

当然，我们一定会问，从何种意义上理解莫妮卡真实地存在在那里？导演杰米•斯图亚特告诉我，她根本不在那里；仅仅是一个幻想植入到大卫的脑海中；结束时的场景完全发生在大卫的脑海里。下载完大卫的全部记忆和信息后，新型机器人再也不需要大卫了。然而就在终结他之前，他们让他的心愿在最后一天得到满足。最后我们得知，他在做梦。那仅仅是大卫的印象。电影前面部分已经确立他既不能睡觉也不会做梦的事实。

一个机器人会关心另一个心愿是否满足了吗？那么，给大卫二十四小时幸福的意义何在呢？如果机器人无法感觉，那么结束场景的意义又在哪里呢？我相信，它意在表明，新型机器人正尝试打造一个它们可以关爱的机器人。它们可以在它们自己的大卫面前扮演妈妈。并且那个机器人会关爱它们。在这样的语境下，爱意味着什么？不多，不少，仅仅是审查、交配、或两者兼有。这就是人工智能的命运。没有妈妈会从此关爱它们。

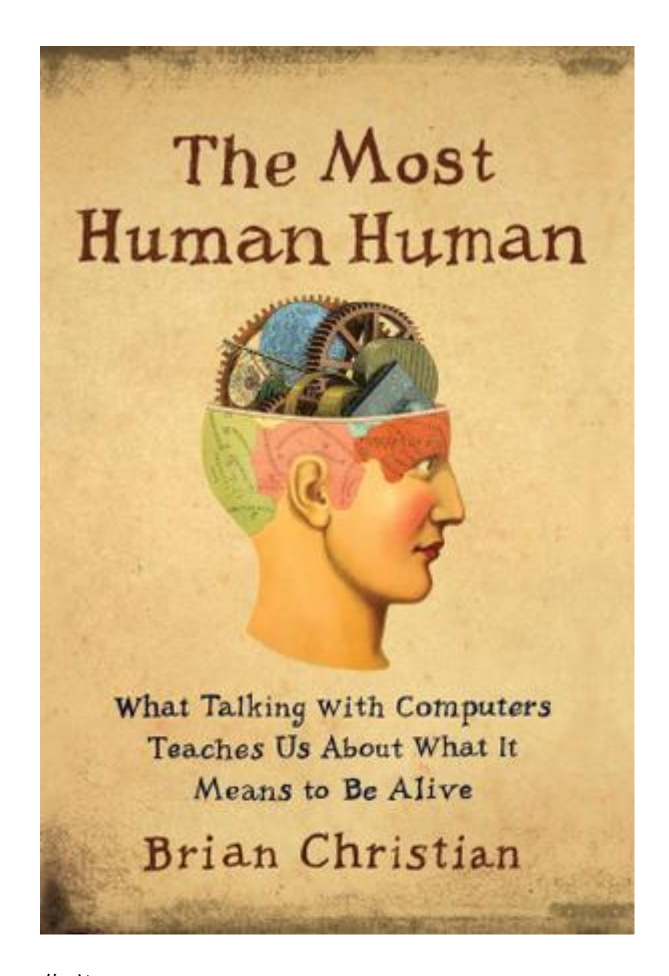
【原文链接】 【回到目录】



【视频】霍金：人工智能对人类可能是致命的

【原文链接】　　【回到目录】

【荐书】The Most Human Human



作者：Brian Christian

副标题：What Talking with Computers Teaches Us About What It Means to Be Alive

出版年：2011-3-1

装帧：Hardcover

ISBN：9780307879158

布莱恩 · 克里斯蒂安（Brian Christian）参加了 2009 年的勒布纳人工智能奖竞赛。克里斯蒂安把自己的参赛经历， 以及他对人、智能、人工智能、人性的思考写成了一本书："The Most Human Human： What Talking With Computers Teaches Us About What It Means To Be Alive"。

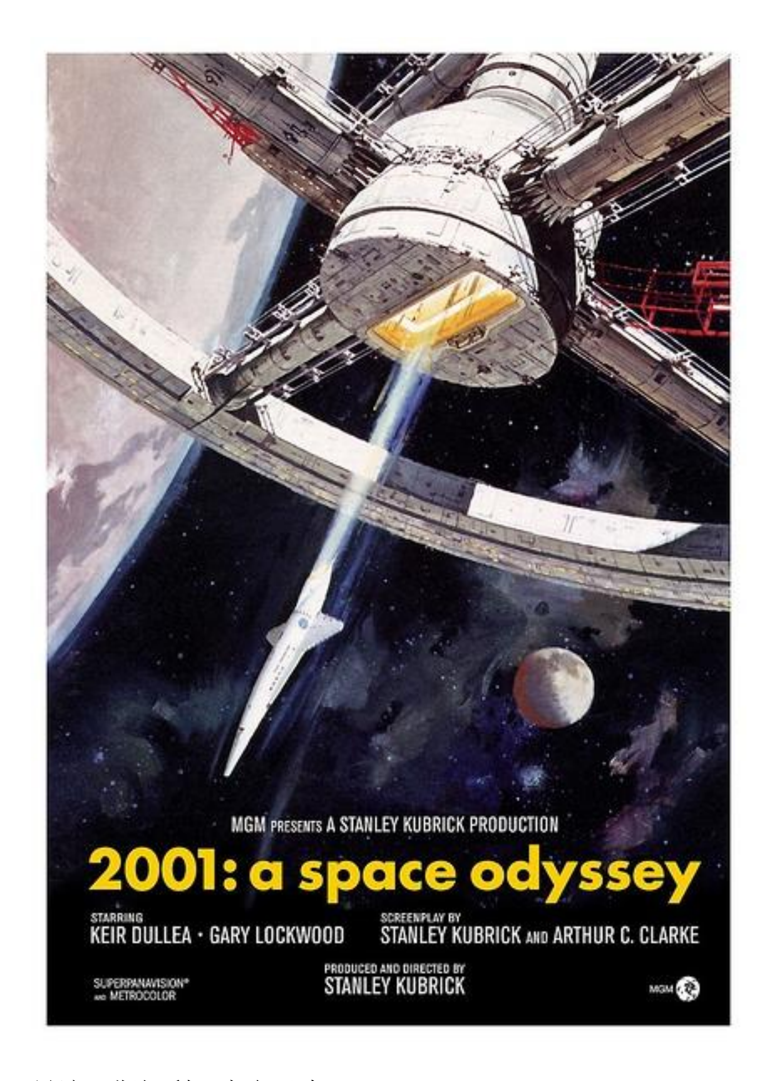


竞赛中， 一位人类裁判向参赛计算机和人类选手提出一系列文字问题，然后根据他们的答案，判断谁是机器、谁是人类。共有 12 位裁判，测试重复 12 次。骗过裁判人数最多的计算机将被授予 “Most Human Computer” 的称号—— “最像人类的计算机” ；相应的，成功说服裁判，让裁判相信自己是人的人类选手，投票最多的那位将被授予 “Most Human Human” 的荣誉

【原文链接】 【回到目录】



【荐影】2001 太空漫游 2001: A Space Odyssey



导演: 斯坦利·库布里克

编剧: 亚瑟·克拉克 / 斯坦利·库布里克

主演: 凯尔·杜拉 / 加里·洛克伍德 / 威廉姆·西尔维斯特 / Daniel Richter / 雷纳德·洛塞特

类型: 科幻 / 悬疑 / 冒险

制片国家/地区: 美国 / 英国

语言: 英语 / 俄语



上映日期: 1968-04-02

片长: 141 分钟 / 160 分钟(premiere cut)

又名: 2001：星际漫游

IMDb 链接: tt0062622

【原文链接】 【回到目录】



主编：张洁平

编辑：施钰涵

设计：潘雯怡，季文仪

校订：童亚琦

出品人：杜婷

若希望订阅此电子周刊 doc 版请发空邮件至 cochinaweeklydoc+subscribe@googlegroups.com；若订阅 pdf 版请发送至 cochinaweeklypdf+subscribe@googlegroups.com；mobi 版至 cochinaweeklymobi+subscribe@googlegroups.com；epub 版至 cochinaweeklyepub+subscribe@googlegroups.com。

此电子周刊由「我在中国」（Co-China）论坛志愿者团队制作，「我在中国」（Co-China）论坛是在香港注册的非牟利团体，论坛理事杜婷、梁文道、闾丘露薇、周保松。除了 Co-China 周刊之外，Co-China 每月还在香港举办论坛，并透过网络进行视频、音频和文字直播。2012 年开始 Co-China 在香港举办面向青年的夏令营，第一届主题为「知识青年，公共参与」，2013 年夏令营的主题是「始于本土：本土、国家、世界冲撞与协商」。

Co-China 论坛网址：https://cochina.co

Co-China 论坛新浪微博：CoChina 論壇（http://weibo.com/1510weekly）

Co-China 论坛 facebook：「我在中國」（Co-China）論壇（https://www.facebook.com/CoChinaOnline）

